

Spare Parts Logistics and Installed Base Information

Muhammad N. Jalil, Rob A. Zuidwijk, Moritz Fleischmann,
and Jo A.E.E. van Nunen

ERIM REPORT SERIES <i>RESEARCH IN MANAGEMENT</i>	
ERIM Report Series reference number	ERS-2009-002-LIS
Publication	January 2009
Number of pages	27
Persistent paper URL	http://hdl.handle.net/1765/14529
Email address corresponding author	mjalil@rsm.nl
Address	Erasmus Research Institute of Management (ERIM) RSM Erasmus University / Erasmus School of Economics Erasmus Universiteit Rotterdam P.O.Box 1738 3000 DR Rotterdam, The Netherlands Phone: + 31 10 408 1182 Fax: + 31 10 408 9640 Email: info@erim.eur.nl Internet: www.erim.eur.nl

Bibliographic data and classifications of all the ERIM reports are also available on the ERIM website:
www.erim.eur.nl

REPORT SERIES
RESEARCH IN MANAGEMENT

ABSTRACT AND KEYWORDS	
Abstract	<p>Many of the challenges in spare parts logistics emerge due to the combination of large service networks, and sporadic/slow-moving demand. Customer heterogeneity and stringent service deadlines entail further challenges. Meanwhile, high revenues rates in service operations motivate companies to invest and optimize the service logistics function. An important aspect of the spare parts logistics function is its ability to support customer-specific requirements with respect to service deadlines. To support customer specific operations, many companies are actively maintaining and utilizing installed base data during forecasting, planning and execution stages. In this paper, we highlight the potential economic value of installed base data for spare parts logistics. We also discuss various data quality issues that are associated with the use of installed base data and show that planning performance depends on the quality dimensions.</p>
Free Keywords	<p>value of information, installed base information, information quality, forecasting, spare parts logistics planning, practice of OR</p>
Availability	<p>The ERIM Report Series is distributed through the following platforms:</p> <p>Academic Repository at Erasmus University (DEAR), DEAR ERIM Series Portal</p> <p>Social Science Research Network (SSRN), SSRN ERIM Series Webpage</p> <p>Research Papers in Economics (REPEC), REPEC ERIM Series Webpage</p>
Classifications	<p>The electronic versions of the papers in the ERIM report Series contain bibliographic metadata by the following classification systems:</p> <p>Library of Congress Classification, (LCC) LCC Webpage</p> <p>Journal of Economic Literature, (JEL), JEL Webpage</p> <p>ACM Computing Classification System CCS Webpage</p> <p>Inspec Classification scheme (ICS), ICS Webpage</p>

Spare Parts Logistics and Installed Base Information

Muhammad N. Jalil*

RSM Erasmus University, Rotterdam, The Netherlands, mjalil@rsm.nl

Rob A. Zuidwijk

RSM Erasmus University, Rotterdam, The Netherlands, rzuidwijk@rsm.nl

Moritz Fleischmann

RSM Erasmus University, Rotterdam, The Netherlands, mfleischmann@rsm.nl

Jo A. E. E. van Nunen

RSM Erasmus University, Rotterdam, The Netherlands, jnunen@rsm.nl

January 2009

Abstract

Many of the challenges in spare parts logistics emerge due to the combination of large service networks, and sporadic/slow-moving demand. Customer heterogeneity and stringent service deadlines entail further challenges. Meanwhile, high revenues rates in service operations motivate companies to invest and optimize the service logistics function.

An important aspect of the spare parts logistics function is its ability to support customer-specific requirements with respect to service deadlines. To support customer specific operations, many companies are actively maintaining and utilizing installed base data during forecasting, planning and execution stages. In this paper, we highlight the potential economic value of installed base data for spare parts logistics. We also discuss various data quality issues that are associated with the use of installed base data and show that planning performance depends on the quality dimensions.

Keywords: Value of Information, Installed Base Information, Information Quality, Forecasting, Spare Parts Logistics Planning, Practice of OR.

* Corresponding Author, Ph. +31 10 40 81414, Address: Room T9-19, P.O. Box 1738, 3000 DR, Rotterdam, The Netherlands.

1. Introduction

The use of information about the customer in the planning and execution of supply chain operations is considered an enabler for better performance towards that customer. This seems particularly valid in the context of spare parts supply chains, where machines are positioned at geographically dispersed customer locations. While facing intermittent or slow moving demand, variety in machine models and part commonality in machines; the positioning of spare parts inventory throughout the geographical network is a critical and delicate task (Cohen et al., 1997; Fortuin et. al, 1999; Huiskonen, 2001).

In a spare parts supply chain, the demand is realized during machine maintenance operation. Accordingly, the data about the realized demand and associated machine's location (i.e., installed base data) may be used for the subsequent planning. Installed base data may consist of machine location data, contractual data, and machine type data. This data can be used to address the challenging task of meeting strict customer deadlines at minimum costs in spare parts supply chains (Oliva et al., 2003; Vigoroso, 2003; Auramo et al., 2005). However, what economic value is generated by the use of installed base data for the planning process still remains to be understood. A major obstacle that companies often face is the issue of varying data quality (Wand et al., 96; Korhonen et al., 1998, Fisher et al., 2000; Lee et al., 2002). Many researchers have also cautioned regarding the potential impact of poor data quality on planning processes (Daganzo, 1984; Bender, 1985). However, the research is scarce for the impact assessment of data error in a real-life spare parts planning situation.

In this paper, we study IBM's spare parts logistics operations to observe the potential economic value of installed base data usage in the real-life spare parts planning. We also analyze the extent to which the attained economic value degenerates by data quality variations in installed base data. IBM spare parts logistics is a frequently cited case in academic literature as an example of the state of the art in spare parts logistics operations (Cohen et al., 1990; Cohen et al., 1997, Fleischmann et al. 2003, Candas et al. 2007; and Kutanoglu, 2008). Similar to other spare parts situations, the primary question that IBM encounters is the placement of spare part inventories in its service network (Cohen et al., 2006). For this purpose, IBM uses a mixed integer program of inventory – distribution optimization logic (see section 2.1 and appendix A). In this paper, we observe the gains of using installed base data for such an optimization program and assess the impact of installed base data quality variations on its performance. In other words, we test the

planning model's robustness to installed base data errors for IBM's spare parts planning environment. The main contributions of this paper are as follows:

1. This study analyzes the benefits of installed base data usage and the detriments of installed base data quality variation on the planning performance by considering scenarios relevant to the spare parts planning at IBM.
2. By using an experimental setting, the paper highlights the role of business environment of a spare parts supply chain that impacts the optimization model's robustness to installed base data errors for the spare parts planning.
3. The paper contributes to the existing body of literature by integrating the data error assessment methods in Operation Research (OR) literature and information quality concepts from Information Sciences (IS).

This paper is organized as follows: First, we describe the important characteristics of spare parts logistics at IBM. We then formulate our research problem to analyze the impact of using machine specific data and its quality on the spare parts planning situation at IBM. In Section 4, we discuss the relevant literature in Operations Research and Information Sciences. In Sections 5 and 6, we depict and discuss numerical results, and we conclude with a discussion on our findings in Section 7.

2. Spare Parts Logistics at IBM

At IBM, the Spare Parts Logistics Organization (IBM-SPO) operates in close coordination with the IBM Service Organization to satisfy the spare parts requirements for after sales service operation. The primary question that IBM-SPO encounters for spare parts planning is how to place the spare parts inventories throughout their service network. For Europe, Middle East, & Africa (EMEA) region, IBM's network consists of almost 150 stock locations. The service network is organized to ensure that IBM has sufficient presence in each region to provide timely service to its business customers. IBM's business customers own IBM's high-end computing machines with various types of service contracts. The service contracts differ in terms of the time to provide complete repair service whenever a failure occurs. The interpreted service deadlines for the delivery of service parts range from 2 hours service deadline to 4 or 8 hrs or even a next day service deadline. Figure 1 indicates the type of field service regions that IBM encounters.

Note the presence of the various customer types in the field service region. Each customer can be served from any of the stock locations that are within the travel distance.

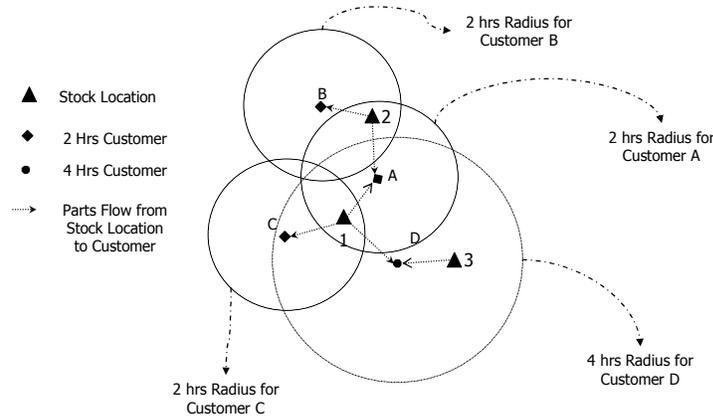


Figure 1. Depiction of IBM's Geographical Service Network for Spare Parts

Consequently for each stock location, the result is the superimposition of more than one service region due to the presence of various customer types. This creates profound complexities for the spare parts planning task.

2.1 Spare Parts Planning

The planning of spare parts inventories is performed by a Mixed Integer Programming based optimization model (Erke et al., 2003) that uses the inventory-distribution type of optimization logic (Thomas et al., 1996). A simplified version of the model that incorporates its main characteristics is depicted in Appendix A. Briefly described, the model follows the optimization logic of Single Echelon – Single Commodity Model (Giani et al., 2004) with additional time based constraints (Candas et al., 2007). The optimization model is a Mixed Integer Program that uses the customer locations, part's unit price, demand rates, travel times, handling and transportation costs as input data. It defines neighbourhood clusters by identifying the stock locations that are within a required service/delivery time for each customer. It seeks to minimize the holding and transportations costs via optimal placement of requisite spare parts inventory in the network while ensuring the service at a given target level to all customers.

The traditional practice in supply chain management is to use stock location level data or POS data for demand inputs. However, it is evident from the above model description, that at IBM, the planning system is designed to use the customer level data as an input. At IBM spare parts logistics, the demand forecasting via extrapolation is impractical at the customer location level due to the slow moving nature of demand. To acquire the demand and location data at the customer level, IBM uses the installed base data as described in the next section.

2.2 Forecasting and Installed Base Information

Machine location information in installed base data is used to derive demand forecasts at customer's zip code level at IBM. In the following example, we describe the demand forecasting procedure at IBM.

In Figure 2a, we depict a geographical region that is serviced by the shown field stock location. The geographical region covers many smaller regions defined as zip codes. The demand forecast (i.e., 20 units) for the stock location has been estimated by applying exponential smoothing to the observed historical demand at the stock location.

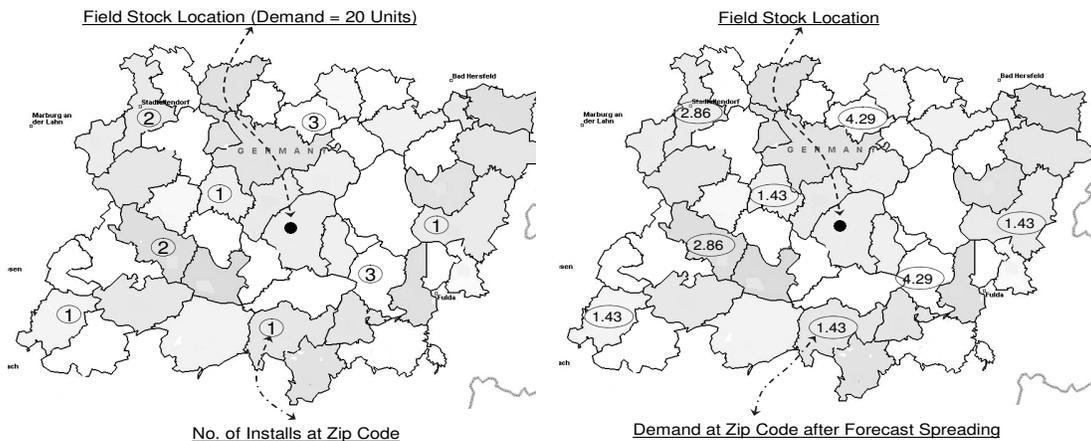


Figure 2a.

Figure 2b.

Figure 2. Forecast Spreading over Installed Base

Figure 2a also depicts the number of installed machines present at nearby zip codes. In Figure 2b, the demand forecast at the stock location is proportionally distributed over these zip codes by taking into account the total number of installs present at each zip code. This procedure facilitates the incorporation of the installed base data to the demand forecasts and provides the demand requirements of each zip code. If the geographical region contains multiple stock locations, then as a first step, the demand forecasts from all stock locations should be accumulated before being proportionally distributed.

2.3 Installed Base Data Quality at IBM

Despite careful considerations, the quality level of installed base data varies due to a number of reasons. For example, the erroneous manual data entry of sales data results in missing data values or wrongly entered data values of installed base data. These errors are homogeneously distributed in all geographical regions. In some situations, the errors are concentrated in a specific

geographical region due to a particular business environment. For example, in some regions IBM sells its machines through a third party sales organization. Due to the procedural mismatch in IT systems of two organizations, inaccurate installed base data is transmitted to IBM. Such types of instances are being classified as heterogeneously distributed errors. In section 6.1, we list various observed errors in installed base data at IBM.

3. Problem Formulation

The availability of geographical information in IBM's installed base data provides an opportunity to use localized demand information at machine's zip code level for the spare parts planning situation described in section 2.1. In such a situation, we attempt to answer the following questions:

1. What additional value is generated by the use of accurate installed base data to derive the machine's zip code level demand forecasts in the spare parts planning optimization?
2. What is the impact of installed base data errors on the outcome of the spare parts planning optimization?

The solution procedure to the first question relies on the comparative analysis of the baseline situation of planning via stock location demand forecasts versus planning via machine location demand forecasts. In the Operations Research (OR) literature, a number of authors have discussed the use of planning optimization models with detailed or disaggregated data vs. aggregated data. However, no conclusive discussion is available in the OR literature for the intended optimization planning context (see section 2.1).

The second question encapsulates the parallel research on data quality in OR and Information Sciences (IS). In the coming section, we review the existing literature on the impact assessment of data quality in OR models. We observe that the notion of data quality in OR literature is inept to encapsulate the various aspects of data quality in a real life spare parts planning situation. Subsequently, we review the data quality in IS literature and map the data quality aspects in the IS literature to the current case. We further analyze the impact of systematic and random data errors on the planning outcomes.

4. Literature Survey

Theoretically, the use of stock location data vs. installed machine's level data can be viewed as disaggregated data usage vs. data aggregation for usage in a planning model. In this section, we review available literature on data aggregation vs. disaggregated data usage for OR planning models. Subsequently, we review the OR literature of data quality assessment and highlight its limitations to accommodate the real life data quality aspects. We then review the data quality notions in IS literature. We also highlight the limitation of data quality assessment procedures in IS to accommodate the decision making context of spare parts planning.

4.1 Data Aggregation vs. Disaggregated Data Usage

The earlier discussion in OR literature for inventory and distribution planning supports the data aggregation due to computational complexity and data acquisition issues (Axsäter 1980; Magee et al. 1985; Rogers et al., 1991; Ballou, 1994; Daganzo, 1996). The researchers attempted to outline the appropriate data aggregation level for transportation, inventory and distribution planning models. Some of these researchers also acknowledged that the data aggregation is a source of potential input error for these planning models. The estimation of value loss due to data aggregation has been highlighted by Ballou (2001) as an unresolved issue for facility location and inventory-distribution class of optimization models.

Owing to the developments in IT sector, organizations are now better equipped to acquire customer data. In a benchmark study, Cohen et al. (1997) discussed the criticality of advance information systems for the design and management of timely service oriented spare parts logistics operations. A survey of spare parts management practices in 310 major companies by Aberdeen group reported that 82% of the companies use ERP or in-house built information systems for spare parts management. In addition, 59% of the companies utilize the state of the art IT systems to capture customer information and machine maintenance history data. The ability to use the detailed customer information in spare parts planning was also stressed (Gecker et al. 2006). Simultaneously, there are developments regarding the computational complexity of the intended class of optimization model (i.e., Single Item – Single Echelon Model). Candas et al. (2007) has shown that the model could be efficiently solved within reasonable timeframes by using advanced heuristics solution methods. However, the potential economic value that could be generated by using detailed data is unaddressed in literature for the intended planning model.

4.2 Data Quality in Operations Research Literature

In the OR modelling literature for supply chain management, two different sources of input data errors are described that may impact the model output (i.e., results): 1) model approximation errors, and 2) data acquisition or sampling errors (Daganzo, 1996). Model approximation errors relate to the modelling assumptions, and approximations made during the modelling process, and data acquisition errors relate to the data which is used as an input to the model. Since, we are interested in the erroneous installed base data; our research questions justify a focus on the data acquisition errors. In a literature survey, Roger et al. (1991) discussed the impact of data aggregation error in Transportation Planning, Multicommodity Distribution Planning, Production Planning and Scheduling problems. Daganzo (1996) studied the impact of data acquisition errors on inventory – distribution problems. The robustness of the Economic Order Quantity (EOQ) type formulation was analyzed with respect to demand data errors. Korhonen et al. (1998) discussed the importance of data accuracy for demand management at Nokia Corporation due to the short product life cycles, and customer retention focus. Cachon et al. (2000) discuss the impact of sampling error on the value of shared information in a two stage supply chain with one supplier and multiple retailers. Toktay et al. (2003) analyze the robustness of various forecasting methods with respect to errors in product return parameters. Thonemann (2001), by using a scenario analysis methodology, shows to what extent the benefits of using advanced demand data in a two-stage supply chain are mitigated due to erroneous demand data.

We note in the above papers that the widely used methodology for error assessment is either the structural analysis of the model or the scenario analysis of the planning model. We also observe that the impact of data errors has not been analyzed for optimization model in this paper. Moreover, in all of the above papers, the definition of quality is limited to the accuracy or sampling error dimension of the data quality. But as we witness in our study, in reality, data quality is a much richer concept than just accuracy or sampling error. To explore the various dimensions of data quality, we review the IS literature on data quality.

4.3 Data Quality in Information Science Literature

The notion of data quality in the IS literature is somewhat different from the OR literature and finds its roots in TQM concepts of quality, i.e., fitness for its purpose. In the IS literature, data quality is defined as a representation of various ontological characteristics of data (Wand et al.,

1996; Mallach, 2000). Some of these ontological aspects of data quality read as follows: *Completeness* indicates to what extent a dataset contains all necessary values: “all values for a certain variable are recorded” (Ballou et al., 1985). *Accuracy* is defined as the degree of agreement between an observed value and an accepted reference value. *Timeliness* is an aspect that characterizes whether the current dataset or data value is out of date for its intended use.

In IS literature, many researchers have attempted to devise a framework to rank the impact of these various quality dimensions (Ballou et al., 1985; Wang et al. 1996; Lee et al., 2002). These survey based studies are limited from an OR application perspective due to the following reasons: First, the surveys were performed mainly on IT professionals who are not the end users of the data. Secondly, the contextual implications of the decision making situation (such as in planning at IBM and the associated geographical nature of errors in IBM’s installed base data; see section 2.3 & 6.1) are not considered during rank assessment.

To enrich the decision making context (i.e., spare parts planning environment) in our study, we utilize a scenario analysis based methodology. First, we analyze the economic value of using installed base data to support spare parts planning. Subsequently, we identify various dimensions of data error in installed base data. In terms of additional costs incurred due to the erroneous installed base data; we analyze impacts for the spare parts planning by using scenario analysis methodology. We conclude the paper with a discussion and interpretation of the results for IBM spare parts planning in particular and spare parts inventory planning in general.

5. Analysis Procedure and Value of Installed Base Data

In this section, we outline the analysis procedure to answer the research questions. To answer the first question, we assess the value of using installed base data by scenario analysis. Figure 3. sequentially depicts the procedure followed for scenario analysis. In the first step, we perform an optimization run by using the optimization model depicted in appendix A for the baseline case with no installed base data incorporation and acquire the baseline stock plan.

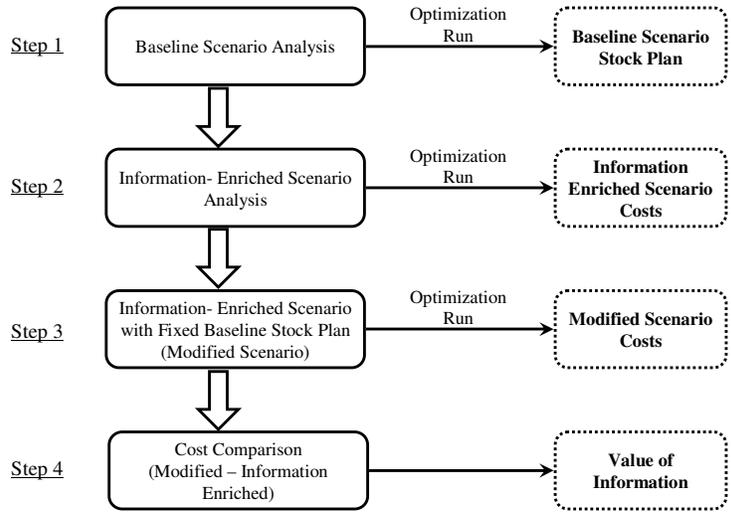


Figure 3. Analysis Procedure

In the second step, we introduce the installed base data (i.e., information enriched scenario) and acquire the costs for the information-enriched scenario by performing the optimization. In the third step, the baseline stock plan is fixed in the information enriched scenario and the costs for this modified scenario are computed. In the fourth step, we compare the costs of modified scenario and information enriched scenario. The difference between the modified scenario and the information enriched scenario indicates the value of that information.

5.1 Incorporation of Installed Base Data in Spare Parts Planning

The installed base data was incorporated into the spare parts planning by using the forecast spreading method described in Section 2.2.

There exists a wide variety in the installed base sizes for various IBM products. For example, a specific machine could have a few hundred customers in the whole EMEA region, whereas a large installed base could have a size of twenty thousand. The impact of changes in installed base size was envisaged and accounted during our analysis by formulating small installed base and large installed base scenarios. The value of installed base data was also analyzed for its sensitivity to demand rates for various installed base sizes. The demand rates and other cost parameters such as transportation costs and holding costs rates are selected according to the observed values of such parameters at IBM.

5.2 Results of Incorporation of Installed Base Data

Scenario 1 - Small Installed Base

By using the procedure outlined in Section 5, we analyze the value of installed base data usage. Figure 4 depicts the results of a small installed base scenario at varying cumulative network demand rates. Due to the slow moving nature of demand, these are the typically observed demand rates at IBM. The installed base size for this test bed scenario is 140 units. The demand rates are in units per week for the complete installed base region. The test bed parameters are listed in Appendix B.

The vertical axis shows the planning costs (i.e., resultant value of objective function; includes transportation and inventory holding costs) of the various scenarios. The information enrichment shows cost improvements (relative percentage changes range from 1% to 16%) as it merges the detailed geographical information about the customer location in the planning process. The gains are considerable for lower demand rates. We argue that in case of lower demand rates, the exact magnitude and positioning of stock units in the network is more critical.

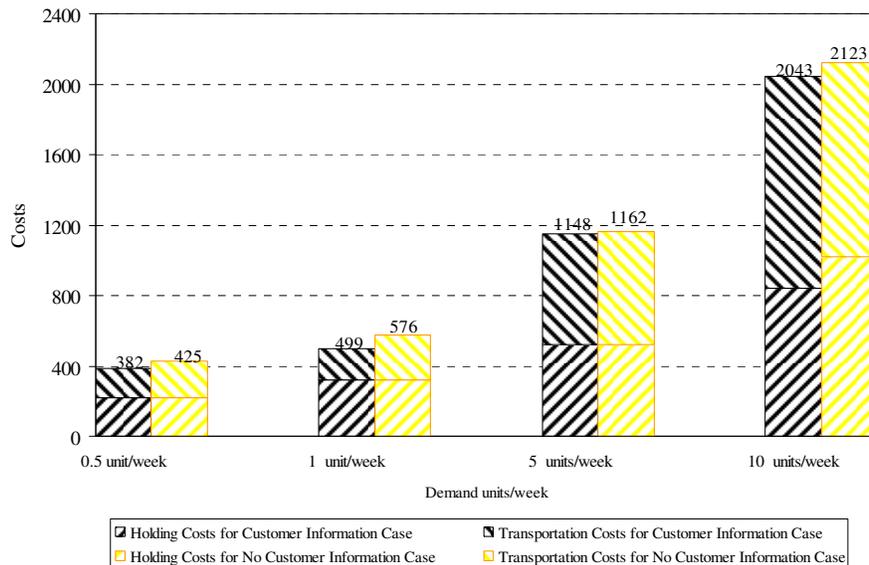


Figure 4. Comparison between No Customer Information and Customer Information Scenario (Information Enrichment)

The use of customer information facilitates the improved stock positioning in the network and subsequently provides cost improvements by reducing the requisite transportation needs. In addition, the installed base also provides the detailed information regarding the neighbourhood region for each customer. For higher demand rate, this customer information led to slightly lower stock requirements. This is due to the fact that less stock units are needed to meet the global

service level constraint during optimization. Due to this, it results in lower inventory holding costs for this scenario.

Scenario 2 - Large Installed Base

In this section, we present the value of installed base data for a large installed base scenario. The size of the installed base is 23,885 machines. For various typically observed cumulative network demand rates (units/week) at IBM, we present the gains of using installed base information in Figure 5.

In this comparative analysis, we observe even higher gains (1% ~ 58%) of using installed base information for various demand rates. The cost savings primarily occurred due to the combination of lower stock requirements and reduction of transportation costs in information enriched scenarios. The costs savings are greater at lower demand rates. The results confirm our earlier argument that the benefits of using installed base information are particularly relevant for lower demand rates.

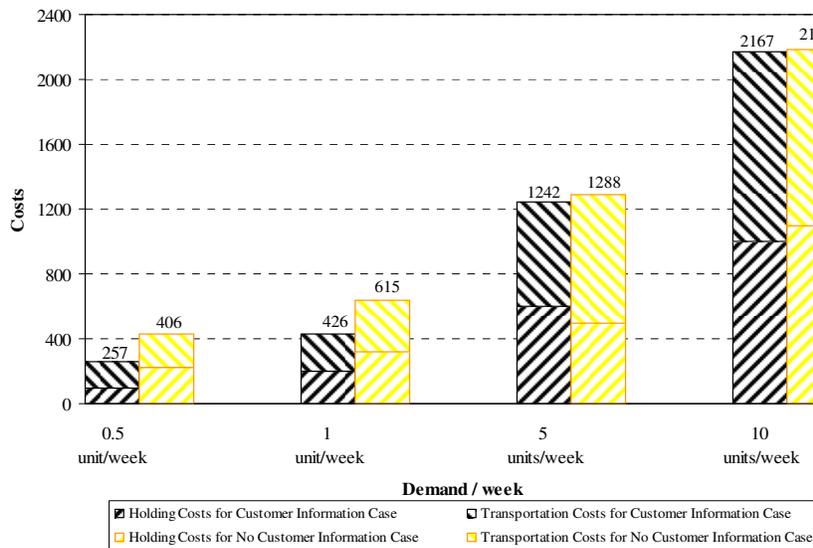


Figure 5. Comparison between No Customer Information and Customer Information Scenario (Information Enrichment)

6 Data Quality and its Analysis

The resultant data quality in a specific database system is the by-product of its design, implementation and usage. It is due to the human-machine interface that most of the erroneous data is generated. This implicitly highlights the need to account for the contextual aspects of the decision situation at hand (e.g. business environment of spare parts planning) during the impact assessment of erroneous data.

6.1 Installed Base Data Quality Variation and its Analysis

There are many systematic reasons that influence the nature of resulting erroneous data. These reasons relate to the business environment of spare parts logistic and are identified via discussions with the planning experts at IBM. In this section, we list some of the reasons that account for frequent data errors in installed base data.

Homogeneously Distributed Error -Missing Value error and Wrong Value Error

Due to human mistakes during data collection and data entry process, we observe completeness and accuracy errors in installed base data. For example, if a specific install is not listed in installed base data, then the completeness aspect of the data quality is observed (i.e., missing value error). On the other hand, if the install is listed, but location information is incorrectly recorded (i.e., wrong value error) then it would be categorized as an error pertaining to the accuracy aspect. In general, these mistakes typically occur at random due to human behaviour during data collection and entry process; therefore, the error is distributed homogeneously throughout the geographical region.

Heterogeneously Distributed Error - Head Quarter Error (i.e. HQ Error)

A common cause for erroneous installed base data could relate to a customer who owns a large percentage of the installed base. In this error scenario, the spare parts logistics provider has listed the customer's company head quarter as machine installation location for all machines. In reality, the machines are installed at many sub-offices of the company. Note that in such a situation, the total size of the installed base remains the same, but a certain number of installs from the complete region are listed at a single location.

Heterogeneously Distributed Error - Primary Stock Location Error (i.e. PSL Error)

In many instances, installed base data contains partial information (e.g. incomplete street or city address) regarding the location of the specific installed machine. In such a case, the planning procedure usually attempts to assign the installed machine to the primary (nearest) stock location.

Heterogeneously Distributed Error - Data Communication Error

In many regions, the OEM sells its machines in partnership with local IT vendors. Due to a mismatch between the data collection procedures and IT systems used, the customer location information is not transferred to the spare parts logistics provider (i.e., Data Communication Error).

6.1.1 Experimental Design

We observe that the errors caused by various business phenomena may have different characteristics in terms of geographical error distribution and data quality dimension. In Table 1, we classify the various completeness and accuracy errors according to the case context of geographical error distribution.

		Case Context – Geographical Distribution	
		Homogeneously Distributed Error	Heterogeneously Distributed Error
Data Quality Characteristics	Accuracy	– Wrong Value Error	– HQ Error – PSL Error
	Completeness	– Missing Value Error	– Data Comm. Error

Table 1. Classification of Error Causes according to their Geographical Distribution

The procedure adopted to analyze these various error scenarios is similar to the procedure outlined in the information enrichment case (see Section 5). In this case, we take the information enriched scenario as a baseline and perform optimization runs to acquire baseline scenario planning costs. Subsequently, we induce errors in the dataset by omitting certain installs according to the specific business phenomenon. We devise various error frequency levels according to each error scenario. For each error frequency, the cost deviations are acquired by fixing the error prone stock plans in the baseline scenario.

In the following sections, we analyze the above completeness and accuracy errors by organizing them according to their geographical distribution context. Table 2 outlines the experimental design for the impact analysis of installed base data quality. As mentioned in Table 2, these error scenarios were analyzed for various demand rates and installed base sizes. The choice of error frequencies is made to understand the impact of data errors on the robustness of planning method at various error concentrations. These error frequencies may not represent the actual error concentration in IBM’s installed base data. Rests of the parameters (such as demand rates and installed base sizes) in these scenarios are formulated to represent the observed characteristic of the planning system. For example, the observation of errors in large installed base sizes is only plausible for homogeneous errors. The choice of associated demand rates for small and large installed base is also in agreement with typically observed demand rates for such installed base sizes.

Scenario		Installed Base Size	Analyzed Error Frequency	Demand Rate
Homogeneous Error Distribution	Completeness Error / Missing Value	Small = 140 Installs	10%, 20%, 30%	1 Unit / week
		Small = 140 Installs	10%, 20%, 30%	10 Unit / week
		Large = 23,885 Installs	10%, 20%, 30%	1 Units / week
		Large = 23,885 Installs	10%, 20%, 30%	10 Units / week
	Accuracy Error / Wrong Value	Small = 140 Installs	10%, 20%, 30%	1 Unit / week
		Small = 140 Installs	10%, 20%, 30%	10 Unit / week
		Large = 23,885 Installs	10%, 20%, 30%	1 Units / week
		Large = 23,885 Installs	10%, 20%, 30%	10 Units / week
Heterogeneous Error Distribution	HQ Error	Small = 140 Installs	10%, 20%, 30%	1 Unit / week
		Small = 140 Installs	10%, 20%, 30%	10 Units / week
	PSL Error	Small = 140 Installs	10%, 20%, 30%	1 Unit / week
		Small = 140 Installs	10%, 20%, 30%	10 Unit / week
	Data Comm. Error	Small = 140 Installs	10%, 20%, 30%	1 Unit / week
		Small = 140 Installs	10%, 20%, 30%	10 Unit / week

Table 2. Experimental Design

6.2 Results of Data Quality Assessment

In this section, we analyze the impact of various information quality aspects outlined in section 6.1. In the first step, we analyze the effect of various error frequencies in the homogeneously distributed error case. Subsequently, we examine the heterogeneously distributed error case by devising the scenarios to depict the underlying business phenomena as outlined in the previous section.

6.2.1 Homogeneously Distributed Error

For the homogeneously distributed error in installed base data, we formulate the associated scenarios for the accuracy and completeness error in small and large installed base size.

Scenario 3: Homogeneously Distributed Completeness Error

Figure 6. depicts the results of the completeness aspect of the homogeneously distributed error (i.e., Missing Value Error) for a small installed base case at demand rates of 1 unit/week (Fig. 6a) and 10 units/week (Fig. 6b). The test bed parameters are listed in Appendix B. We observe that

despite having a highly erroneous installed base data, the results show trivial losses due to data quality degradations.

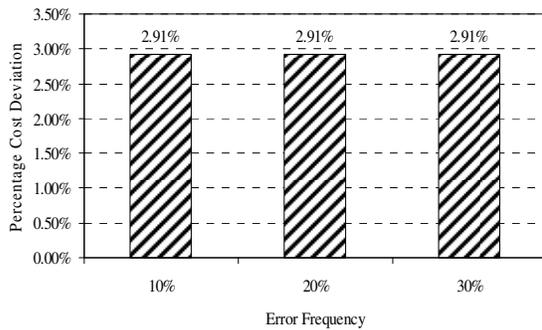


Figure 6a. Demand 1 unit/week

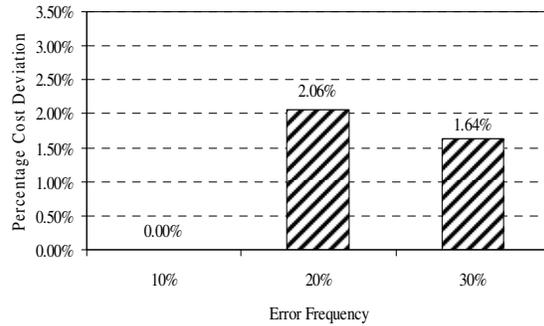


Figure 6b. Demand 10 units/week

Figure 6. Percentage Cost Deviation from Baseline for Completeness Aspect

The losses can be explained from the fact that due to homogeneously induced error, the demand sizes at each individual zip code homogeneously change throughout the network. As a result, the overall geographical distribution of installed base is preserved to a large extent. In erroneous scenarios, the requisite quantity of stocks in the entire network is similar to the baseline case. The only variation in costs results due to the slight geographical shifts in stock positioning decisions. This incurs cost variations in terms of additional transportation costs. For lower demand rates, these variations are unable to offset the gains of using installed base information. In case of higher demand rate (i.e. 10 units/week), we observed in Section 5.2, that the gains of using installed base data are small. Therefore, the small losses in higher demand rate (Figure 6b) are able to negate the gains of using installed base data.

The above argument regarding preservation of geographical distribution for homogeneously induced error is particularly valid in large installed base scenarios. These scenarios were analyzed according to the listed variations in experimental design section (Table 2). It should be noted that no losses were observed for large installed base scenarios at all error frequencies and demand rates. We argue that in large installed base, we tend to normalize any asymmetric effects of errors in the geographical distribution of installs.

Scenario 4: Homogeneously Distributed Accuracy Error

In this scenario, we simulate the homogeneously distributed accuracy dimension due to data entry errors. In this case, the erroneous install listing represents the installed machine which is listed at the wrong address in installed base data. The test bed parameters are listed in Appendix B.

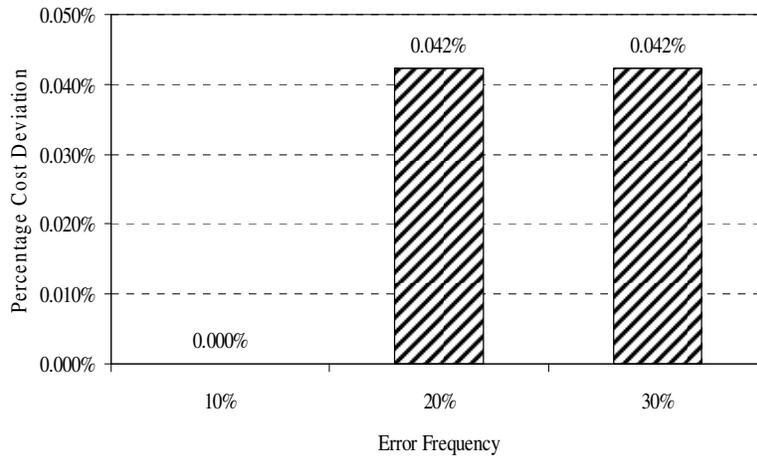


Figure 7. Demand 1 unit/week

Figure 7. Percentage Cost Deviation from Baseline for Accuracy Aspect

Figure 7 depicts the results of this experiment for small installed base at the demand rate of 1 unit/week. We observe that the negligible impact of erroneous installed base data. The explanation for the low impact is similar to the previous case of completeness error. The cost deviations are incurred due to the additional transportation costs. For 10 units/week demand rate scenarios of small and large installed base, and 1 unit/week demand rate for large installed base. We do not observe any impact of error scenarios in our analysis.

We observed in the previous scenarios for the homogeneous error that despite having very high frequency of errors in installed base, the losses are somewhat insignificant compared to the gains of using installed base. This is particularly valid for lower demand rates, where additional information considerably improves the stock positioning decision. In the next section, we analyze the heterogeneously distributed errors in installed base data to observe their impact on spare parts planning.

6.2.2 Heterogeneously Distributed Error

In this section, we analyze the impact of heterogeneously erroneous data due to various business situations listed in Section 6.1. We formulate these scenarios by using the information enriched situation as a baseline and induce errors in the dataset by following the underlying behaviour of the business phenomena.

Scenario 5: Heterogeneously Distributed Accuracy Error - HQ Error

Due to the specific business situation (as outlined in Section 6.1), the concentration of installs is listed at a single location; whereas in reality, these installs are geographically dispersed at various locations. We formulate this behaviour at varying levels of error intensity and subsequently observe the system behaviour at different demand levels. The test bed parameters are again listed in Appendix B. Figure 8 shows the planning system behaviour for this scenario.

Due to the slow moving nature of spare parts demand, 1 unit per week (for the complete network) is the typically observed demand for such installed base size. The 10% error case relates to a customer who owns 10% of the total installs. The spare parts logistics provider has listed all the specific customer's installs at a company's head quarter location. In reality, the machines are dispersed homogeneously throughout the geographical network at the sub-locations.

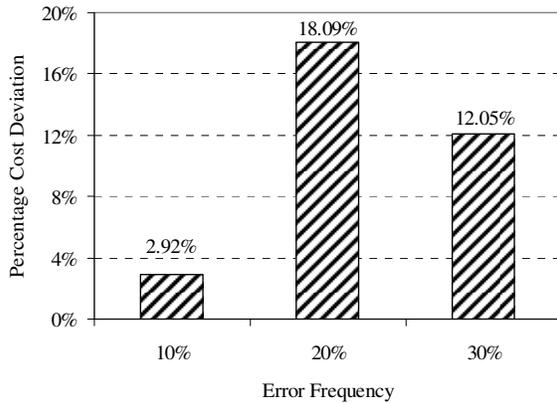


Figure 8a. Demand: 1 unit/week

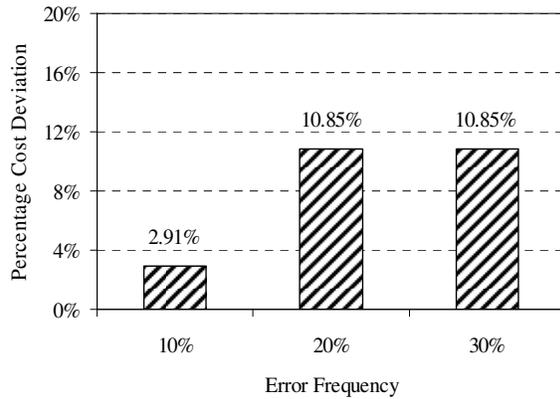


Figure 8b. Demand: 10 units/week

Figure 8. Percentage Cost Deviation from Baseline (HQ Error)

We observe cost variations from 2.92% ~ 18.09% approximately for various error cases at 1 unit/week demand (Figure 8a). For higher demand category (i.e., 10 units/week, Figure 8b), the cost deviations are 2.91% ~ 10.85%. Note that in both of these scenarios; the costs of poor data quality offset the benefits of using installed base data for some error frequencies (see Section 5.2).

For lower demand rate scenario, the losses are observed due to the following situation. During the optimization, the stock outflows originating from a single stock location to all nearby customer locations are accumulated. The corresponding stocking decision for the stock location is the lowest integer value that is higher than the accumulated outflows from that stock location. Due to the HQ error, we place higher demand requirement at a single zip code. Therefore, the

optimization accumulates all these flows to the nearby stock location and satisfies the service level constraints at a lower stock placement level. As a result, the overall stock placement in the network is lower than the baseline case due to the accumulated flows at a single stock location. There is a higher build-up of stock at a single location. Therefore, the transportation costs to serve the actual installed base (i.e., baseline installed base) are also higher.

In higher demand case, as a result of optimization, we tend to place higher stock units at the HQ error inducing customer's zip code. This results in extra holding and transportation costs due to erroneous placement of stocks at a specific location.

Scenario 6: Heterogeneously Distributed Accuracy Error - PSL Scenario

In this scenario, the installs with incomplete location data are assigned to the nearest stock location's zip code. We analyze this scenario by using the test bed parameters listed in appendix B. Figure 9 depicts the results for this scenario. For demand rate 1 unit/week we observe the cost variations of 0.00% ~ 1.64%. In this case, the total stock placement in the entire network is same for baseline and erroneous scenario. The minor cost deviations relate to the increased transportation costs due to inaccurate stock positioning in the erroneous scenarios. Similar is the situation for the higher demand rate, where additional transportation costs result for some error frequencies due to the inaccurate positioning of stock.

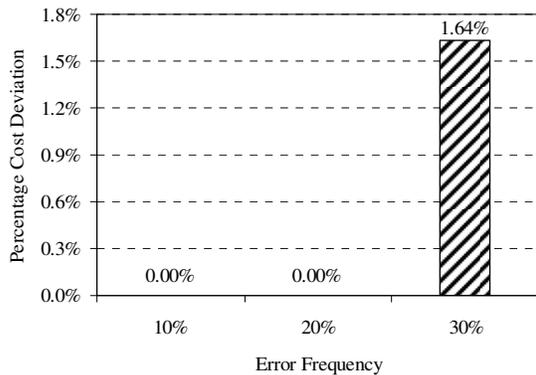


Figure 9a. Demand 1 unit/week

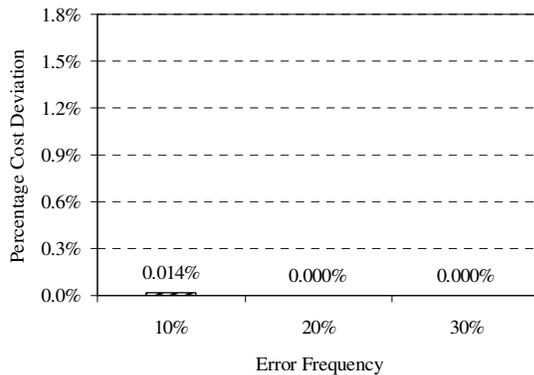


Figure 9b. Demand 10 unit/week

Figure 9. Percentage Cost Deviation from Baseline (PSL Error)

Scenario 7: Heterogeneously Distributed Completeness Error - Data Comm. Error

In many situations, OEM sells its products through a partnership with the local vendor. Due to IT infrastructural or procedural shortfalls, the sales information is not fully communicated to the

OEM. This results in erroneous installed base for a specific sub-region. In this section, we analyze this situation and observe the cost variations for such scenario at various demand rates.

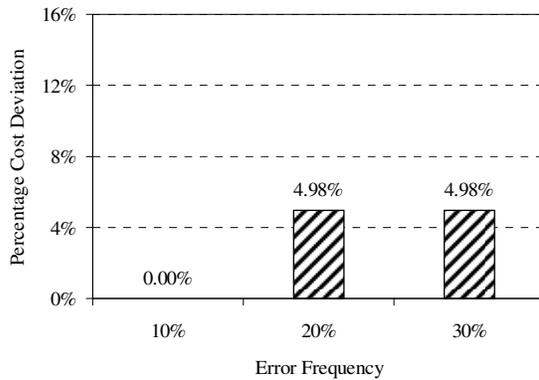


Figure 10a. Demand 1 unit/week

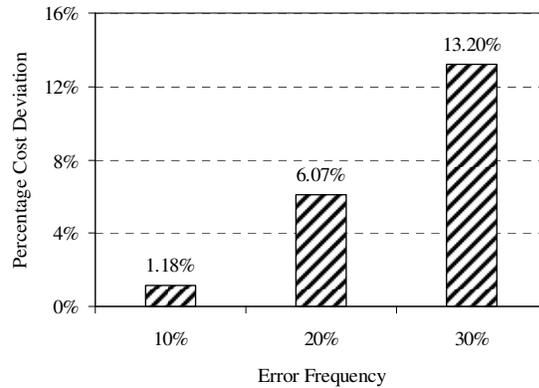


Figure 10b. Demand 10 units/week

Figure 10. Percentage Cost Deviation from Baseline (Data Comm. Error)

Figure 10a depicts the relative cost deviations of 0.00% ~ 4.98% for 1 unit/week demand rate. The inaccurate placement of stocks in erroneous scenarios leads to the increase in transportation costs. Figure 10b depicts the results for the higher demand rate. In this case, the combination of inaccurate stock quantities and inaccurate stock positioning contributes to the losses. Also note for this scenario, that the losses due to inaccuracies in installed base are much higher than the observed gains of using installed base data.

We observe in the above heterogeneous error scenarios that there are additional costs incurred due to the variations in installed base data quality. Depending on the error structure induced by each of the error scenario, the results vary. In many cases, the losses due to inaccuracies negate the benefits of using installed base data.

7. Discussion & Conclusions

In this paper, we analyzed the gains of using installed base data in spare parts planning. We also identified the various types of data errors that are present in installed base data, and analyzed the impact of these errors on spare parts planning performance. Table 3 summarizes the results of the numerical study. The positives represent the situations where the overall value enhancement due to the information enrichment is not deteriorated by the data quality errors. The negatives represent the situations where data quality errors have negated the value enhancement of information enrichment. We observed that the gains of using installed base data are significant.

This is particularly relevant for low demand rates (a predominant characteristic in spare parts logistics).

Installed Base Size	Demand Rate	Information Enrichment with No Error	Information Enrichment with Error				
			Homogeneous Error		Heterogeneous Error		
			Missing Value Error	Wrong Value Error	HQ Error	PSL Error	Data Comm. Error
Large	Low	+	+	+	N/A	N/A	N/A
	High	+	+	+	N/A	N/A	N/A
Small	Low	+	+	+	-	+	+
	High	+	-	+	-	+	-

Table 3. Summary - Effects of Information Enrichment

We classify the various types of data errors in installed base data as homogeneous and heterogeneous with respect to the geographical distribution of machines. We observe in homogeneous error scenarios that the large frequencies of errors in installed base data typically do not result in significant impact. Therefore, the benefit attained by information usage is preserved. This is related to the robustness of the planning method. We observe that the planning method positions inventories in the network based on the geographical distribution of demand. If the overall geographical distribution of demand is preserved, the planning method shows little sensitivity to the geographical displacement of customer locations. In the homogeneous error case, despite having large errors percentages; the regional demand's contribution to optimization remains the same. Because, by having errors homogeneously distributed, we tend to homogeneously inflate the demand for the rest of the installs by proportionally distributing demand over the installed base. Due to this, we do not observe significant cost deviations in homogeneous error cases. Observe that we assume that the size of the installed base is accurate in all scenarios. The allows us to compare the scenarios using costs analyses.

In general, we observe higher cost deviations in heterogeneous error cases. The geographical synergy effect of installed base spreading and planning method does not provide robustness against the heterogeneous errors. Since, by having error heterogeneously distributed, we tend to disturb the proportionality of the original geographical distribution of demand. For the heterogeneous error scenarios, we observed that the gains of installed base data usage deteriorated.

To summarize, we list the following implications and conclusions from our analysis.

1. There are potentially significant gains in using detailed customer's geographical information (i.e., installed base data) for the planning of network inventory in a multiple stock locations and geographically dispersed customers setting.
2. It is beneficial to use installed base data for spare parts planning. However, one should identify, understand and align the business environment of data acquisition and usage to acquire maximum gain.
3. Data quality has been researched from various standpoints. In OR research, It has been mainly analyzed from sampling error aspect, whereas, in many OR application situations, the realistic situation extends to various other quality dimensions. Therefore, it is useful to extend the data quality research in OR to include the data quality dimensions as described by IS research. However, one should account for the planning context. For example, we observe that accuracy error is present in both homogeneously and heterogeneously distributed error cases. Only in the heterogeneous case, the impact is significant.
4. We also argue that spare parts planning managers should attempt to understand the business phenomena that induce certain data errors. This supports the prioritization of data quality improvement investments.
5. We also observe that the relative gains are greater for smaller demand rates, since the stock positioning decision is more important due to less available redundancies in that case. Therefore, there is a strong case for using customer's geographical information for slow moving demand situations.
6. We should caution however, that the results of the numerical study should not be over-generalized. The study was performed on a single echelon – single commodity type of geographical stock allocation model with time based constraints. The geographical overlapping of serving locations and time based constraints play an important role in our findings. The same results may not completely hold if we study a model that does not allow the overlapping serving locations. Besides spare parts placement in spare parts supply chains, many facility location and network design problems are similar in nature. The results of this study may also be applicable in those settings.

Acknowledgement

This research is partially sponsored by Transumo ECO project number GL05022b. The authors also acknowledge the contributions of IBM, The Netherlands to this project. The views presented in this paper represent the opinion of the authors only and does not necessarily represent the viewpoints or policy of IBM, The Netherlands.

References

- Auramo J and Ala-Risku T (2005). Challenges for Doing Downstream. *International Journal of Logistics: Research and Applications* 8:4 333-345.
- Axsäter S (1981). Aggregation of Product Data for Hierarchical Production Planning. *Opns Res* 29:4 744-756.
- Ballou D and Pazer H (1985). Modelling Data and Process Quality in Multi-Input, Multi-Output Information Systems. *Mngt Sci* 31:2 150-162.
- Ballou R (1994). Measuring Transport Costing Error in Customer Aggregation for facility Location. *Transp J* 33:3 49-59.
- Ballou R (2001). Unresolved Issues in Supply Chain Network Design. *Information Systems Frontiers* 3:4 417-426.
- Bender P S (1985). Logistics System Design. In: Robeson J F and House R G (eds). *The Distribution Handbook 1985*. The Free Press: New York.
- Cachon G P and Fisher M (2000). Supply Chain Inventory Management and the Value of Shared Information. *Mngt Sci* 46:8 1032–1048.
- Candas M F and Kutanoglu E (2007). Benefits of Considering Inventory in Service Parts Logistics Network Design Problems with Time-Based Service Constraints. *IIE Trans* 39:2 159–176.
- Cohen M A, Kamesam P, Kleindorfer P R, Lee H L, and Tekerian A (1990). OPTIMIZER: IBM's Multi-Echelon Inventory System for Managing Service Logistics. *Interfaces* 20:1 65–82.
- Cohen M A, Zheng Y and Agrawal V (1997). Service Parts Logistics: A Benchmark Analysis. *IIE Trans* 29:8 627-639.
- Cohen M A, Agrawal N and Agrawal V (2006). Winning in the Aftermarket. *Harvard Bus Rev* 84:5 129-138.
- Daganzo C (1987). Increasing Model Precision Can Reduce Accuracy. *Transp Sci* 21:2 100:105.

Daganzo C (1996). *Logistics Systems Analysis*. Springer-Verlag.

Erke Y, Christina Ma Y H and Booth M C (2003). Method of Determining Inventory Levels. *United States Patent Publication* 0061126A1.

Fisher M L, Raman A and McClelland A S (2000). Rocket Science Retailing is Almost Here: Are You Ready. *Harvard Bus Rev* 78:4 115-124.

Fleischmann M, van Nunen J A E E and Grave B (2003). Integrating Closed-Loop Supply Chains and Spare Parts Management at IBM. *Interfaces* 33:6 44–56.

Fortuin L and Martin H (1999). Control of Service Parts, *Int J Oper Prod Mngt* 19:9 950-971.

-Geoffrion A M and Graves G W (1974). Multi-commodity Distribution System Design by Bender Decomposition. *Mngt Sci* 20:5 822-844.

Ghiani G, Laporte G and Musmanno R (2004). *Introduction to Logistics Systems Planning and Control*. John Wiley and Sons.

Huiskonen J (2001). Maintenance Spare Parts Logistics, Special Characteristics and Strategic Choices. *Int J Prod Econ* 71: 125-133.

Korhonen P, Huttunen K and Eloratna E (1998). Demand Chain Management in a Global Enterprise – Information Management View, *Prod Plann Control* 9:6 526-531.

Kutanoglu E (2008). Insights into Inventory Sharing in Service Parts Logistics Systems with Time-Based Service Levels. *Comput Ind Eng* 54:3 341-358.

Lee W L, Strong D M, Kahn B K and Wang R Y (2002). AIMQ: A Methodology for Information Quality Assessment. *Information and Management* 40: 133-146.

Magee J F, Copacino W C and Rosenfield D B (1985). *Modern Logistics Management: Integrating Marketing Manufacturing and Physical Distribution*. John Wiley & Sons: New York.

Mallach E G (2000). *Decision Support and Data Warehouse Systems*. McGraw-Hill,.

Oliva R and Kallenberg R (2003). Managing the Transition from Products to Services, *Int J of Service Industry Management* 14:2 160-172.

Rogers D F, Plante R D, Wong R T and Evans J R, (1991). Aggregation and Disaggregation Techniques and Methodology in Optimization. *Opns Res* 39:4 553-582.

Thomas D J and Griffin P M (1996). Coordinated Supply Chain Management, *Eur J Opl Res* 94:1 1-15.

Thonemann U W (2002). Improving Supply Chain Performance by Sharing Advance Demand Information. *Eur J Opl Res* 142:1 81-107.

Toktay B, van der Laan E and de Brito M P (2003). Managing Product Returns: The Role of Forecasting. Working Paper, ERIM Report Series, Erasmus University, Rotterdam, The Netherlands.

Vigoroso M (2003). *Service Parts Management: Unlocking Value and Profit in the Service Chain*. Aberdeen Group.

Gecker R and Vigoroso M (2006). *Best Practices in Service Chain Performance Management: Measuring the Value of a Profit-Centric Service Operation*. Aberdeen Group.

Wand Y and Wang R (1996). Anchoring Data Quality Dimensions in Ontological Foundations. *Communication of the ACM* 39:1 86-95.

Wang R Y and Strong D M (1006). Beyond Accuracy: What Data Quality Means to Data Consumers. *J of Management Information System* 12:4 5-33.

Appendix A

Parameters

H	Holding costs per unit
T_{jl}	= 1 whenever travel time between customer j and location l within pre-specified time, = 0 otherwise
V_{jl}	unit transportation and handling costs between customer j and location l
D_j	demand rate customer j
α	required service level

Sets

L	set of locations
J	set of customer(s) (regions)

Decision variables

S_l	Stock level at location l
λ_{jl}	Flow rate from sending location l to customers j

Objective function to be minimized (inventory holding and transportation costs)

$$\sum_{l \in L} HS_l + \sum_{j \in J} \sum_{l \in L} V_{jl} \lambda_{jl}$$

Global service level constraint (demand delivered in time)

$$\sum_{j \in J} \sum_{l \in L} T_{jl} \lambda_{jl} \geq \alpha \sum_{j \in J} D_j$$

Demand of each customer must be satisfied

$$\sum_{l \in L} \lambda_{jl} = D_j \quad \forall j \in J$$

Outflow from a stock location must not exceed than the units present in stock

$$\sum_{j \in J} \lambda_{jl} \leq S_l \quad \forall l \in L$$

$$S_l \in \mathbf{Z}^+ \quad , \quad \lambda_{jl} \geq 0$$

Appendix B

Test Bed Parameters

Scenario No.	Normal Transportation Cost Rate from Field Locations	Emergency Transportation Cost from Central Location	Holding Cost Rate for Field Locations	Holding Cost Rate for Central Location	Unit Cost	Service Level
1	Fixed: 57, Var.: 0.12/km	1250	0.25	0.05	400	0.9
2	Fixed: 57, Var.: 0.12/km	1250	0.25	0.05	400	0.9
3	Fixed: 57, Var.: 0.12/km	1250	0.25	0.05	400	0.9
4	Fixed: 57, Var.: 0.12/km	1250	0.25	0.05	400	0.9
5	Fixed: 57, Var.: 0.12/km	1250	0.25	0.05	400	0.9
6	Fixed: 57, Var.: 0.12/km	1250	0.25	0.05	400	0.9
7	Fixed: 57, Var.: 0.12/km	1250	0.25	0.05	400	0.9

Table 4. Test Bed Parameters

Publications in the Report Series Research* in Management

ERIM Research Program: “Business Processes, Logistics and Information Systems”

2009

How to Normalize Co-Occurrence Data? An Analysis of Some Well-Known Similarity Measures

Nees Jan van Eck and Ludo Waltman

ERS-2009-001-LIS

<http://hdl.handle.net/1765/14528>

Spare Parts Logistics and Installed Base Information

Muhammad N. Jalil, Rob A. Zuidwijk, Moritz Fleischmann, and Jo A.E.E. van Nunen

ERS-2009-002-LIS

<http://hdl.handle.net/1765/14529>

* A complete overview of the ERIM Report Series Research in Management:
<https://ep.eur.nl/handle/1765/1>

ERIM Research Programs:
LIS Business Processes, Logistics and Information Systems
ORG Organizing for Performance
MKT Marketing
F&A Finance and Accounting
STR Strategy and Entrepreneurship