TUCS

Tomas Eklund

The Self-Organizing Map in Financial Benchmarking

Turku Centre *for* Computer Science

TUCS Dissertations
No 56, October 2004

The Self-Organizing Map in Financial Benchmarking

Tomas Eklund

DOCTORAL DISSERTATION

To be presented with the permission of the Faculty of Economics and Social Sciences at Åbo Akademi University for public criticism in DataCity, Entrance A, third floor, Auditorium 3102 DataCity, on the 10th of December 2004, at 12 o'clock noon.

Åbo Akademi University
Turku Centre for Computer Science /
Institute for Advanced Management Systems Research
Lemminkäinengatan 14B, FIN-2520 Åbo
Finland

Åbo, 2004

SUPERVISOR

Professor Barbro Back
Institute for Advanced Management Systems Research
Department of Information Systems
Åbo Akademi University

REVIEWERS

Professor Teija Laitinen
Department of Accounting and Finance
University of Vaasa

Professor Olli Simula
Department of Computer Science and Engineering
Helsinki University of Technology

OPPONENT

Professor Jaakko Hollmén
Department of Computer Science and Engineering
Helsinki University of Technology

ISBN 952-12-1436-8
ISSN 1239-1883

Åbo Akademis Tryckeri
Turku, Finland
2004

ABSTRACT

Managers in today's international companies are facing a very complex competitive environment. In an increasingly global market, it has become more and more important to recognize who your competitors are, what they are doing, and how they are performing. This is the objective of competitor analysis.

Financial benchmarking is an essential component of competitor analysis. Financial benchmarking is performance benchmarking using financial measures. However, while today's tools are adequate for measuring performance according to a few ratios for a small number of competitors, increasing dimensionality and data present difficulties for managers. Analyzing ratios that measure different aspects of performance is especially difficult using today's tools. A state of the art survey of financial benchmarking methods in Finnish publicly-noted companies found that few advanced, multiple ratio methods are currently used. The survey found support for the need for new, complexity-reducing tools in financial benchmarking.

In this dissertation, a model for financial benchmarking in the international pulp and paper industry has been built using the self-organizing map (SOM). Financial data for 98 companies in the international pulp and paper industry have been collected for the years 1995-2002, and seven financial ratios measuring different aspects of financial performance were calculated. A SOM model has been built, and the benchmarking of the companies has been tested. The model has been evaluated by subject matter experts from industry.

The study has found that the managers considered the model better than many of their own methods, especially in terms of format. In particular, the model was found to be useful in strategic decision making settings. The results thus indicate that the SOM is a feasible tool for financial benchmarking, and presents several advantages over current methods.

Keywords: Financial benchmarking, state of the art, pulp and paper industry, self-organizing maps, user evaluation

ACKNOWLEDGEMENTS

For many doctoral students, the dissertation process is a long and lonely path to pursue. For my part, it has indeed been challenging, but never lonely, thanks to the many people who have supported me in the process.

First and foremost, I would like to thank my supervisor, Professor Barbro Back. Barbro has always had the ability to see the big picture, and to keep me focused on a goal that could sometimes be quite difficult for me to see. Also, Barbro's open door policy has provided me with endless support and encouragement, which is extremely important for any doctoral student. No matter how busy she has been, she has always found time to answer my frequent questions. I do not think that I could have had a better supervisor for my thesis.

I would also like to thank Professor Christer Carlsson and the entire Department of Information Systems and IAMSR at Åbo Akademi University. Christer makes it a point to always have encouraging and inspiring comments for his doctoral students, and always seeks to motivate them to greater feats. I am also grateful to Professor Pirkko Walden for her comments and shared experience, in particular when I was constructing the survey instruments for this research. The administrative staff, Leena, Sirpa, and Stina, has always been helpful and kind, and have helped to provide an inspiring working environment.

I am extremely grateful to Professor Hannu Vanharanta, Tampere University of Technology, Pori, and Professor Ari Visa, Tampere University of Technology. Hannu has been a never-ending source of encouragement and help. He has provided me with many opportunities to develop and apply my research, and is a never tiring project planner. Hannu is always enthusiastic and encouraging, and is a continuing source of support for me. Ari has provided me with the technological support that a business major needs in order to apply the SOM in a business application. Especially preprocessing issues have been challenging, and this is an area in which Ari has helped me a lot.

I sincerely thank Professor Teija Laitinen from the University of Vaasa and Professor Olli Simula from the Helsinki University of Technology, for functioning as my reviewers. Their comments have helped me to considerably improve the overall quality of this thesis. I would also like to thank Professor Jaakko Hollmén for accepting the invitation to function as my opponent at the public examination.

I am very grateful to the members of the GILTA research team, Aapo Länsiluoto, Antonina Durfee, Antti Arppe, Camilla Magnusson, Jarmo Toivonen, Jarno Mäkinen, Jonas Karlsson, and Tomi Vesanen. In particular, Antonina has always encouraged and supported me as a colleague and friend, and has been a recurring co-author. Aapo, another friend and co-author, has shared many of my frustrations with aspects of this research, but it was his drive and ambition that finally lead to the joint evaluation of our results. Working with Jonas during our master's and licentiate theses was very fruitful, and without our spirit of teamwork and friendly competition, reaching this stage would not have been possible.

I would like to thank the entire staff of TUCS for providing me with the financial and administrative support that most doctoral students can only dream of. The travel grants, publication series', cross-university courses, and not least, the granted income, have allowed me to concentrate my full-time efforts on my research. Precious few doctoral students in ours or any other fields enjoy the luxury of this level of support, and therefore, I am very grateful. Many other institutions have also supported this work financially. I gratefully acknowledge the support of the Academy of Finland (DOMINO, grant no. 104639), the Finnish Technology Center, Tekes (GILTA, Grant No. 40943/99, and 4M grant no. 40373/03). I also gratefully acknowledge the grants provided by Willy Gielens stipendiefond at Åbo Akademi University.

One of the things I have enjoyed in particular during this process has been the opportunity to meet with industry representatives and to apply my methods in direct cooperation with companies. I sincerely thank Metso Corporation, and in particular Markku Karlsson, Eeva Mäkelä, and Marko Palmunen, for the opportunity to work with Metso. I also gratefully acknowledge the many companies and managers who took part in our surveys and research projects.

I wish to sincerely thank Professor Bo-Christer Björk from the Swedish School of Economics and Business Administration, who together with Professor Olli Simula, functioned as my opponent at my licentiate seminar. Their valuable views and comments have helped me on the road to this thesis.

Professor Unni Lindström and Professor Katie Eriksson, Åbo Akademi University, have always taken a personal interest in the progress of my research, and I greatly value their support and views on doctoral studies and research from the perspective of a different field of research.

I have also had the opportunity to work with many great co-authors. For this, I would like to thank Antti Arppe, Barbro Back, Mikael Collan, Adrian Costea, Antonina Durfee, Jonas Karlsson, Aapo Länsiluoto, Camilla Magnusson, Hannu Vanharanta, and Ari Visa. Stefan Lindström deserves my thanks for his help in the coding of the online survey instrument, and Tero Lehto for his help in the encoding of the answers. I have also enjoyed working with everyone in the Data Mining and Knowledge Management laboratory at TUCS. In particular, Piia

Hirkman has helped me to take the first steps towards a testable prototype of the system, and is my guide to the world of Extreme Programming.

I would like to thank my many friends at IAMSR who have provided me with many hours of stimulating discussions. I will always remember our (sometimes very long) coffee and lunch breaks during which we debated a wide range of different topics, from science to religion, from technology to social problems. These have been a valuable counterweight to my research. I would especially like to thank Niko Herlin, who always found time to arrange a recreational evening for the rest of us. My friends from my studies here in Turku also deserve my gratitude for making this period in my life so memorable.

Last but not least, my thanks go to my family. My parents, Birgitta and Gösta Eklund, have always supported me in my choices and taught me the value of critical thought and analysis. Finally, I would like to thank my wife, Ulrika Gustafsson, for being there for me and supporting me in my decisions. This dissertation process, as any, has at times been taxing, and there have been many long weekends and evenings apart. Thank you, Ika, for your patience, love, and dedication.

Turku, October 21, 2004

TABLE OF CONTENTS

LIST OF PUBLICATIONS

1. Eklund, T., B. Back, H. Vanharanta and A. Visa (2004). "Financial Benchmarking Tools in Finnish Companies – A State of the Art Survey." TUCS Technical Report No. 618, Turku Centre for Computer Science, Turku.

2. Eklund, T., B. Back, H. Vanharanta and A. Visa (2003). "Financial Benchmarking Using Self-Organizing Maps – Studying the International Pulp and Paper Industry." Data Mining - Opportunities and Challenges. J. Wang. Hershey, PA, Idea Group Publishing: 323-349.

3. Eklund, T., B. Back, H. Vanharanta and A. Visa (2003). "Using the self-organizing map as a visualization tool in financial benchmarking." Information Visualization **2**(3): 171-181.

4. Länsiluoto, A., T. Eklund, B. Back, H. Vanharanta and A. Visa (2004). "Industry Specific Cycles and Companies' Financial Performance - Comparison with Self-Organizing Maps." Benchmarking: An International Journal **11**(4): 267-286.

5. Kloptchenko, A., T. Eklund, B. Back, J. Karlsson, H. Vanharanta and A. Visa (2004). "Combining data and text mining techniques for analyzing financial reports." International Journal of Intelligent Systems in Accounting, Finance and Management **12**(1): 29-41.

6. Magnusson, C., Arppe, A., Eklund, T., Back, B., Vanharanta, H. and Visa, A. (200x). "The language of quarterly reports as an indicator of change in the company's financial status." Accepted for publication in Information & Management (forthcoming).

7. Eklund, T., B. Back, H. Vanharanta, and A. Visa (200x). "Evaluating a SOM-Based Financial Benchmarking Tool." Submitted to the International Journal of Accounting Information Systems (in review process).

LIST OF FIGURES AND TABLES

PART I
Research Summary

1 INTRODUCTION

1.1 Background

There are many parties interested in the financial performance of a company. Investors want to find promising investments among the thousands of stocks available on the market today. Managers want to be able to compare the performance of their own company to that of others, in order to isolate areas in which the company could improve. Creditors want to analyze the company's long-term payment ability, and auditors want to assess the reliability of a company's financial statements. Financial analysts want to compare the performance of a company to that of others, in order to find financial trends on the markets. A tool commonly used by these parties is financial competitor benchmarking. (Bendell et al. 1998)

Benchmarking involves comparing various aspects of a company's operations to those of another, in order to discover areas in which the company could improve. Benchmarking can be divided into two groups, *qualitative* and *quantitative*, depending upon the data being used in the comparison. Qualitative benchmarking involves using descriptive data, which describe various aspects of a company's activities. Quantitative benchmarking, on the other hand, involves using numerical data, usually financial information. Quantitative benchmarking is often referred to as financial benchmarking, since it often involves using financial ratios calculated based on the financial performance of a company. (Karlöf 1997)

Financial comparisons of this kind are commonly published in financial journals, for example, *Talouselämä*, *The Financial Times*, and *Business Week*. Similar comparisons are also routinely published in industry-specific journals, such as *Paper and Timber Magazine* and *Pulp and Paper International*, for the pulp and paper industry. Although these comparisons can be considered to be examples of financial benchmarking, they are usually rather limited in the number of financial ratios used. For example, *Pulp and Paper International's* annual Top 150 list of the largest pulp and paper producing companies in the world (Rhiannon et al. 2001), is based on net sales, and no other information is used. This kind of information is next to worthless to investors, as it is impossible to make any long-term decisions based solely on these rankings.

Instead, many investors would rather compare more of the information found in the companies' annual reports. However, the problem with these comparisons is that the amount of data gathered quickly becomes unmanageable. Using ordinary spreadsheet programs, one can easily compare two to six companies at a time according to one ratio at a time. However, if one wants to obtain an overview of the competitors on the market, or want to take into account several ratios at the same time, spreadsheet programs are difficult to use. For example, comparing two ratios for ten companies for a period of five years produces 100 ratios to compare. Comparing five ratios for the same amount of companies during the same period produces 250 ratios. The availability of financial information, through the Internet, is also now greater than ever before. This leads to a situation, faced by many managers and investors today, in which the amount of data available greatly exceeds the capacity to analyze it (Adriaans and Zantinge 1996).

Managers, decision makers, analysts, and stakeholders could basically use spreadsheets and graphs for benchmarking purposes, and are currently doing so. However, the problems mentioned above restrict the feasibility of using these simple tools for large investment decisions. This problem is illustrated in Table 1.1, where a benchmarking of financial ratios for five companies is illustrated for one year. The ratios each reflect different aspects of financial performance. For example, operating margin is a profitability ratio that measures a company's ability to create a profit on its sales. Equity to capital, on the other hand, measures the solvency (i.e. long-term payment ability) of a company. From the table, it is easy to deduce that analysis of this type of multidimensional data is difficult without some kind of visualization tool, as natural order can not be used to determine the "best performer". The analysis becomes even more tedious as additional years of data are introduced.

Company in 2001	*Operating Margin*	*ROE*	*ROTA*	*Equity to Capital*	*Quick Ratio*	*Interest Coverage*	*Receivables Turnover*
Stora Enso	11.01	10.55	7.68	42.93	0.83	4.19	6.23
Georgia-Pacific	3.14	-7.66	2.74	17.14	0.51	0.79	9.89
International Paper	-1.27	-10.79	-0.85	25.97	1.04	-0.59	8.67
Smurfit-Stone Container Corp	7.44	3.07	5.89	22.66	0.63	1.40	13.39
Oji Paper	5.79	2.91	2.11	25.48	0.54	2.55	3.94

Table 1.1. Multiple ratio analysis using spreadsheets.

A possible solution to this problem is to use data-mining tools. Data-mining tools are applications used to find hidden relationships in data. Data-mining tools offer the advantages of large scale, multidimensional comparisons, in less time and with error handling capabilities. In addition, some data-mining tools have strong visualization capabilities. One data-mining tool that could be particularly suitable for the problem in this case is the self-organizing map.

Self-organizing maps are two-layer neural networks, which use the unsupervised learning method. Self-organizing maps group data according to patterns found in the dataset, making them ideal tools for data exploration. Tasks suitable for self-organizing maps include clustering, categorization, visualization, information compression, and hidden factor analysis.

1.2 Aim of the research

The aim of this research project is to build and evaluate a model for financial benchmarking in the international pulp and paper industry using self-organizing maps. A database consisting of calculated financial ratios for a number of pulp and paper companies, for the period 1995-02, will be created. The source of these data will be the individual companies' annual reports. A data-mining tool, specifically the Kohonen self-organizing map (Kohonen 1997), will be used to analyze the data found in the database. A financial benchmarking of the companies will be performed. The results of the benchmarking will be compared to existing domain knowledge in the form of the textual parts of the annual reports, as well as industry publications, in order to determine the model's fidelity with real-world phenomena. The model will be evaluated by a number of subject matter experts.

1.3 Related research

There have been several studies on self-organizing maps in the past. According to Oja et al. (2003), 5,384 papers concerning Kohonen's self-organizing maps had been published by 2001. However, although many papers on self-organizing maps have been published, very few studies have dealt with the use of self-organizing maps in financial benchmarking.

Most applications of self-organizing maps have dealt with speech recognition, engineering applications, mathematical problems, and data processing (Kaski et al. 1998). Some examples of research papers include cloud classification (Ambroise et al. 2000), image object classification (Becanovic 2000), breast cancer diagnosis (Chen et al. 2000a), user profiling of mobile phone users for fraud detection (Hollmén et al. 1999; Hollmén 2000), document collection analysis and organization (Honkela et al. 1997; Kohonen et al. 2000; Lagus 2000), and extracting knowledge from text documents (Visa et al. 2000).

As was mentioned above, the self-organizing map has been widely applied to engineering problems. For example, Tryba and Goser (1991) applied the self-organizing map to monitor the state of a distillation process. The process was monitored according to a number of input and output measures, such as temperature and input volume. The authors conclude that the SOM could be used for automatic process control. This research is further explored in Simula et al. (1996), and Alhoniemi et al. (1999). The authors of the later study applied the

SOM to industrial process monitoring and modeling, with cases from the pulp and paper and steel industries. In the paper, the applications included the monitoring of a continuous pulp digester, the modeling of a steel production process (Himberg et al. 2001), and technology analysis of different pulp and paper mills from around the world (Simula et al. 1999a). Simula et al. (1999b) include time-series prediction in addition to the above.

Vesanto (2002) developed a complete framework for the exploration of data using the self-organizing map. The author has also explored a number of different visualization techniques.

Kaski and Kohonen (1996) and Kaski (1997) used the self-organizing map to create a map illustrating the international distribution of welfare and poverty. The map created in the study was based on 39 welfare indicators, and clustered several countries depending upon these indicators. Kaski and Kohonen (1996) also noted that this approach could be applicable to the financial grading of companies as well.

An example of the application of neural networks for financial analysis is the study by Martín-del-Brío and Serrano-Cinca (1993). Martín-del-Brío and Serrano-Cinca used self-organizing neural networks for two different financial applications. The first application was an attempt to predict bankruptcies among Spanish banks during the 1977-85 banking crisis. The map was based on nine financial ratios, with the main focus on solvency. The study included 66 banks. Of these banks, 29 eventually went bankrupt. Although the network was unable to correctly predict all of the bankruptcies, it was able to predict most outcomes, and the map was clearly divided into the two areas bankrupt and solvent. The result is supported by a number of similar studies using neural networks utilizing supervised learning.

The second application was a self-organizing map for the study of the financial state of Spanish companies. In this study, five financial ratios were used. 84 companies, for the period 1990-91, were included. The study concentrated on the solvency of the companies. A number of clusters of companies were identified, and the changes from 1990 to 1991 were noted. One of the findings was that of an overall decrease in the profitability of Spanish companies during 1991.

Serrano-Cinca (1996) continued the study by building a complete decision support system for financial diagnosis based on the self-organizing map, complemented with both multivariate statistical models, such as linear discriminate analysis (LDA), as well as with neural models, such as the multiplayer perceptron network (MLP).

Deboeck and Kohonen (1998) explore a number of financial applications of the SOM. These include interest rate projections, mutual fund selection, corporate failure analysis, and real-estate valuation.

Back et al. (1995) and Kiviluoto (1998) also explored use of the SOM for bankruptcy prediction. Back et al. (1995) created three models for bankruptcy prediction. The models were based on a backpropagation network, a SOM, and a Boltzmann Machine. The backpropagation model was found to outperform the other models. Kiviluoto (1998) used the SOM in a supervised manner, providing the outcome (healthy or bankrupt) as a variable, in addition to a number of financial indicators. In this manner, an output map indicating the bankruptcy risk space was created. The author thereafter built a classifier based on a SOM-based radial basis function (RBF) network. The results of the experiment were encouraging, with the RBF-SOM classifier performing better than a number of other techniques.

Another similar study is by Tan et al. (2002), who used the self-organizing map for credit rating. The authors calculated the 18 financial ratios used by Standard and Poor's Corporation (S&P) for 300 US companies in the Consumer Cyclicals sector. They then trained a SOM based on these data, leaving out variables (ratios) that provided no additional value to the discriminating ability of the map. The objective of the study was to see if the SOM could be used to provide the same classification as the credit companies achieve using their methods. Although the study relied on only quantitative measures (S&P uses a combination of quantitative and qualitative measures), the results did come close to the actual ratings of the companies. The conclusion of the study was that the SOM could be used to create a credit rating classification model.

Besides the studies by Martín-del-Brío and Serrano-Cinca (1993), Serrano-Cinca (1996), and Tan et al. (2002), there are two previous studies that relate strongly to this thesis. In the first study, Back et al. (1998b), the authors compared 120 companies in the international pulp and paper industry. The study was based on standardized financial statements for the years 1985-89, found in the Green Gold Financial Reports database (Salonen and Vanharanta 1990a; 1990b; 1990c). The companies used in the experiment were all based in one of three regions: North America, Northern Europe or Central Europe. The objective of the study was to investigate the potential of using self-organizing maps in the process of investigating large amounts of financial data. Back et al. (2001) is a follow-up study to the 1998 paper, with the addition of textual analysis.

The SOM has been widely studied and tested in the literature. For example, Mangiameli et al. (1996) compared the SOM with seven popular hierarchical clustering methods: *single linkage*, *complete linkage*, *average linkage*, *centroid method*, *Ward's method*, *two-stage density*, and *Kth nearest neighbor*. The authors compared the tools using 252 datasets with various levels of imperfections, including *data dispersion, outliers, irrelevant variables,* and *nonuniform cluster densities*. The SOM was found to be "superior to all seven hierarchical clustering algorithms commonly used today" and a better tool for "decisions that require the cluster analysis of messy data such as market segmentation, credit analysis, quality problems, and operations problems" (Mangiameli et al. 1996). Kiang and Kumar (2001) made a comparison between self-organizing maps and factor analysis and K-means clustering. The authors compared the tool's performances on simulated data, with known underlying factor and cluster structures. The results of the study indicate that self-organizing maps can be a robust alternative to traditional clustering methods. Wang and Wang (2002) conclude that there are three main advantages associated with using SOM over K-means clustering. Firstly, SOM does not rely on any assumptions of statistical tests; basically, there are few other assumptions than the dataset itself (Wang 2001). Secondly, the SOM can better deal with data that do not have regular multivariate distributions than statistical clustering methods can. Finally, visualization of the results is a very strong feature of the SOM. Based on these conclusions, the SOM will be used for this experiment.

1.4 Overview of the dissertation

In this chapter, an introduction to this thesis has been provided. The aim of the research has been defined, and finally, related research has also been presented.

In Chapter 2, the research methodology used in this thesis is presented. The thesis will use the constructive research approach. Evaluation of the results is critical to the constructive research approach. Therefore, Chapter 2 also contains a review of information systems success measures, and motivation for the choice of evaluation measures used in this thesis.

Chapter 3 presents the current state of the art in financial benchmarking methods in Finnish publicly-noted companies. Firstly, the key concepts and potential competing techniques are presented. Then, the results of an expert survey of Finnish financial benchmarking methods are presented. The study seeks to determine the current state of the art, as well as manager's satisfaction with current methods.

In Chapter 4, the concept of benchmarking is defined and explained. The type of benchmarking used in this thesis, financial competitor benchmarking for performance analysis, is presented.

In Chapter 5 financial ratio analysis (FRA) is discussed. Firstly different classification patterns for financial ratios are presented. Then, accounting differences that can affect the outcome of the experiment are discussed. Ways of reducing the effect of accounting differences are presented, and the effect of accounting differences on individual financial ratios is shown. International Accounting Standards (IAS), and their effects, are also discussed. Finally, the ratios used in this thesis are presented, and the selection is motivated.

In Chapter 6, neural networks, specifically self-organizing maps, are discussed. Firstly, the background and definition of neural networks are given. Then, the self-organizing map is presented in detail.

In Chapter 7, the companies included and the basis for their selection will be presented. The data collection process, and the handling of missing or incomplete data, will also be discussed.

In Chapter 8, the training and analysis of the financial benchmarking model are presented. Firstly, the properties of the data, and the required preprocessing, are discussed. Then, the training process is illustrated. Identification of the clusters on the map and a benchmarking of the Top 5 international pulp and paper companies are performed. Finally, interesting extensions, multilevel analysis and combined quantitative-qualitative analysis, are discussed.

In Chapter 9, a subject matter expert (SME) evaluation of the benchmarking model is presented, and the results and conclusions are discussed.

Finally, in Chapter 10, conclusions and suggestions for further research are presented.

An overview of the thesis is shown in Table 1.2.

Topic	**Part I**	**Part II**
Introduction	Chapter 1	
Research framework	Chapter 2	
Key concepts, current state of the art and managers' satisfaction with current methods	Chapter 3	*Papers* 1, 7
Benchmarking	Chapter 4	
Financial Ratio Analysis	Chapter 5	
Neural Networks and Self-organizing maps	Chapter 6	*Paper* 2
The selected set of companies	Chapter 7	
Training and analysis of the model, benchmarking the Top 5 companies, multilevel analysis, combining quantitative and qualitative analysis	Chapter 8	*Papers* 2, 3, 4, 5, 6
Expert evaluation of the financial benchmarking model	Chapter 9	*Paper* 7
Conclusions and future research	Chapter 10	

Table 1.2. Overview of the thesis.

1.5 Contributions and publications

The primary contribution of this thesis is an end/business user validated financial benchmarking model for the pulp and paper industry. Firstly, the thesis establishes the current state of the art in financial benchmarking in Finnish, publicly-noted companies, and illustrates the potential for more advanced tools. The thesis then illustrates the technical construction and application of a financial benchmarking model using the SOM, based on generally accepted guidelines in the literature. Conclusions concerning the technical construction, such as preprocessing issues, are made. Finally, the model is evaluated by a number of subject matter experts in a face validation setting. The results show that the model was perceived as valid by the managers. Based on the results achieved, a number of conclusions and recommendations for building a prototype for more extensive testing are presented.

In the literature, there have been a number of efforts to validate the SOM or individual SOM models, primarily using technical measures (e.g. Flexer 2001; Kiang and Kumar 2001; de Bodt et al. 2002). Vesanto (2002, p.69) suggested user evaluations as a validation approach. However, user evaluations, especially end/business user evaluations, have not been widely used in the literature. However, as Vesanto (2002, p.69) states, the validity of the results of explorative approaches is ultimately determined by the end user. The results of explorative approaches, such as the SOM, are difficult to validate using technical measures. In this thesis, such an evaluation is performed in order to determine the usability of the tool for business use. The result is a validated model, as well as valuable information for the construction of a testable prototype, a topic of future research.

During the course of the incremental research, the results have been reported in a number of scientific papers.

Paper 1, “Financial Benchmarking Tools in Finnish Companies – A State of the Art Survey”, is a state of the art survey of Finnish publicly-noted companies. The contribution of the paper is to establish the current state of the art in financial benchmarking methods in Finnish companies. The study shows that few advanced methods are currently used, and provides support for the potential added value of the SOM. It has been published as a technical report in the Turku Centre for Computer Science Technical Reports publications series. I was the main author of the paper.

Paper 2, “Financial Benchmarking Using Self-Organizing Maps – Studying the International Pulp and Paper Industry”, introduces the concept of financial benchmarking using self-organizing maps. It provides a general framework for the approach, and introduces a number of different benchmarks that can be performed. It combines much of what was learned during the initial phases of my research. The publication has been published as a chapter in an edited book, *Data Mining – Opportunities and Challenges*, and has been blind peer-reviewed before acceptance. I was the main author of the paper.

Paper 3, "Using the self-organizing map as a visualization tool in financial benchmarking", represents incremental improvements over Paper 2. The main contribution of the publication is that it presents the final model that is being evaluated in the thesis. It also introduces a number of important changes to the framework in Paper 2, such as improved preprocessing methods and updated data, as well as corrects a number of errors in the original experiment. Finally, while Paper 2 focused on the SOM as a data-mining approach, Paper 3 adds more focus on the visualization capabilities of the SOM, something very central to this thesis. Paper 3 has been published in a blind peer-reviewed international journal, *Information Visualization*. Although it is again a joint paper, I was the main author.

Paper 4, "Industry Specific Cycles and Companies' Financial Performance - Comparison with Self-Organizing Maps", discusses the importance of multilevel environment analysis in financial benchmarking. The primary contribution of the paper is that it seeks to explain factors that influence individual companies' financial performance through the combination of two benchmarking models: an industry level model and a firm level model. Although further research in the area is required, the paper shows the importance of this type of simultaneous analysis, as some explanatory factors are found. Paper 4 has been published in a blind peer-reviewed international journal, *Benchmarking – An International Journal*. It is a joint paper in which I was primarily responsible for the training and analysis of the firm-level model. The analysis of the combined results was carried out as a joint effort between the two main authors.

Paper 5, "Combining data and text mining techniques for analyzing financial reports", suggests and explores the combination of quantitative and qualitative techniques for financial analysis. Although it uses a very limited database, the paper's main contribution lies in showing the potential value of such analysis. It clearly shows the importance of further research in this area, as well as the potential for such tools. Paper 5 has been published in a blind peer-reviewed international journal, *The International Journal of Intelligent Systems in Accounting, Finance and Management*. It is a joint paper in which I was primarily responsible for the analysis and updating of the SOM model. The analysis of the combined results was carried out as a joint effort between the two main authors.

Paper 6, "The language of quarterly reports as an indicator of change in the company's financial status", explores the use of a linguistic method, specifically collocation networks, as a complement to the SOM in financial analysis.

Collocation networks provide a potential tool for combined quantitative/qualitative analysis of financial information. The paper has been accepted for publication in a blind peer-reviewed international journal, *Information & Management*. It is a joint paper in which I was primarily responsible for the analysis and updating of the SOM model. The analysis of the combined results was carried out as a joint effort between the three main authors.

Paper 7, “An Expert Evaluation of the SOM in Financial Benchmarking in the International Pulp and Paper Industry”, describes the subject matter expert evaluation of the final SOM benchmarking model. The paper is a joint paper in which I am the main author. The paper has been submitted to a blind peer-reviewed international journal, the *International Journal of Accounting Information Systems*.

2 RESEARCH FRAMEWORK

This chapter will address the research framework of this thesis. Firstly, general research assumptions, and the classification of approaches in social sciences and information systems research, are discussed. The constructive research (design science) approach is presented. Then, the research approach in this thesis is explained. Finally, evaluation of information systems is discussed.

2.1 Research Approach

Burrell and Morgan (1979) divide scientific research approaches[1] in social sciences into two dimensions: *objective* and *subjective*. The authors' argument is based on differences in four assumptions: *ontological*, *epistemological*, *human nature*, and *methodological assumptions*.

The ontological assumptions imply how the researcher perceives reality: is reality of an objective nature (*realism*), i.e. equal for all, or is it subjective (*nominalism*), i.e. "the product of one's mind" (Burrell and Morgan 1979). Realism can also be divided into two schools: *naïve realism* and *scientific realism* (Goles and Hirschheim 2000). Naïve realists view the universe as being comprised of objectively given, immutable objects and structures, and thus is independent of the observer's appreciation of them. Scientific (or *critical*) realism holds that although reality is objective and independent of the user's appreciation of it, reality can only be understood through the use of non-immutable models. Nominalists (or *idealists*) see reality as a subjective construction of the mind, which is therefore influenced by socially transmitted concepts such as languages and cultures.

The epistemological assumptions, *positivism* and *anti-positivism*, refer to the nature and grounds of knowledge, how and if it can be transferred, and especially to the validity and proof of knowledge. Positivism seeks to explain reality by finding regularities and relationships between elements, i.e. by deriving laws and theorems. Positivism has its roots in traditional natural sciences. Anti-positivists, on the other hand, claim that the world can only be understood from the view of individuals who are directly involved in the studied activities. Anti-positivists

[1] From here on, general *approaches* are taken to consist of the combined assumptions and acceptable methods of the researcher, *paradigms* represent the generally accepted assumptions within a specific scientific community, *methods* are the individual tools (surveys, interviews, laboratory experiments, etc.) available to the researcher, and *research methodologies* constitute the methods valid under the assumptions of a particular paradigm or approach.

reject the notion that a researcher can understand human activities by being entirely objective, and claim that no entirely objective knowledge can be gained from science.

The human nature assumptions deal with the relationship between human beings and their environment. The two extremes on this level are the *deterministic* and *voluntaristic* viewpoints. According to determinism, the individual's actions are determined by the environment and external circumstances, i.e. responding mechanically to changes in their environment, whereas voluntarism ascribes humans a stronger free will and an ability to change and affect their environment.

Based upon these assumptions, the choice of methodology falls into one of two classes, either *nomothetic* or *ideographic*. Nomothetic methodologies are based upon systematic protocol and technique. They are based on the methods and approaches used in the natural sciences, and follow the pattern of quantitative testing of hypotheses according to scientific rigor. Surveys and questionnaires are typical of this approach. Ideographic methodologies stress the importance of "getting close to" and obtaining first hand knowledge of the object of research. Iivari (1991) adds the *constructive research methodology* to the methodology assumptions of Burrell and Morgan's classification framework. The author claims that the special nature of IT research as applied science requires an own set of methods. These are *conceptual development*, the developing of conceptual models and frameworks that do not describe any existing reality, but strive to create a new one, and *technical development*, which produces physical artifacts, e.g. software.

One of Burrell and Morgan's (1979) conclusions was that paradigms are incommensurable, i.e. it is not possible to work within the assumptions of two paradigms at the same time (Goles and Hirschheim 2000). This has further fueled the ongoing paradigm wars, with proponents of a particular paradigm primarily taking the stance of either positivism or anti-positivism. While positivism has been very successful in the natural sciences, it is often considered too restrictive for social sciences.

Hirscheim and Klein (1989) apply and adapt Burrell and Morgan's framework to classify research in the area of information systems development. They conclude that there is no universal paradigm in this area of research. However, Goles and Hirscheim (2000) and Mingers (2001) note that the vast majority of IS studies are *functionalist* (objective, positivist) in their assumptions. Goles and Hirscheim call for more interpretive research in the field of IS, and Mingers argues for a pluralist approach to IT research.

Galliers (1992) provides another classification model for research methods, but refers to the approaches as *scientific* (or *empirical*) and *interpretivist approaches*.

Scientific (or *empirical*) approaches have arisen from scientific tradition, stressing repeatability, reductionalism, and refutability. These approaches are based on, for example, laboratory and field experiments, surveys, and case studies. Scientific approaches match Burrell and Morgan's (1979) nomothetic methodologies (objective dimension), while the interpretivist approaches correspond to Burrell and Morgan's (1979) ideographic methodologies (subjective dimension). The model by Burrell and Morgan provides a general framework for dividing social sciences into two dimensions, under which most approaches can be classified. Galliers concentrates on the methods themselves, as well as providing a discussion of the advantages and drawbacks with each of these.

2.1.1. Constructive Research or Design Science

Research can be divided into *basic* and *applied research* (Simon 1969). Basic research is concerned with identifying what is a part of reality, and creating a model, theory, or framework that describes this. Applied research is concerned with applying what has been learned through basic research. Research in IS, as in engineering, is often applied research, although the classification is determined by the intent of the research (March and Smith 1995). Research can also be divided into *descriptive* and *prescriptive research*. Natural science is descriptive, i.e. an effort to understand the object of study, and design science is prescriptive (or normative), i.e. attempting to improve performance (March and Smith 1995).

As was mentioned earlier, IS research is often applied research, which requires a different research approach than is traditional within the natural sciences. The objective of applied research is to produce research that is applicable in the real world (Galliers and Land 1987). One approach that is gaining in popularity is the *constructive approach*. The constructive approach deals with the creation of entities (e.g. models, diagrams, plans, etc.) that produce solutions to explicit problems, and their usability can be demonstrated through the actual implementation of the solution (Kasanen et al. 1993). In this sense, the constructive approach is normative, i.e. prescriptive, instead of descriptive, as natural sciences tend to be. The value of the construct must also be demonstrated by proving that it offers a better solution to a problem than the previous one did (Järvinen 2001, p.101). The constructive approach can produce both conceptual and technical artifacts (Iivari 1991). Constructive research is typically based on existing research knowledge and/or new technological advances (Järvinen 2001; Kasanen et al. 1993).

The constructive approach[2] is inherently an interpretative approach, essentially positioned in the subjective dimension of Burrell and Morgan's (1979) framework. According to the constructive approach, constructs are used to make sense of the real world, and the context and perspective are therefore considered to determine the ontological "truth", thus making the assumptions nominalistic (idealistic) and anti-positivistic according to the Burrell and Morgan framework (Iivari et al. 1998; Lincoln and Guba 2000, p.168). As an interpretive approach, the constructive approach is usually associated with case studies and qualitative methods, but quantitative methods are also commonly used (Kasanen et al. 1993).

The constructive methodology was pioneered by Simon (1969), Iivari (1991), Kasanen et al (1993), and March and Smith (1995). March and Smith presented a research framework based on the inherent nature of IT research. This framework is shown in Figure 2.1. The framework divides IT research activities into two main categories: *design science*[3] and *natural science*. According to the authors, natural science deals with explaining natural phenomena, and answering questions like how and why. Natural sciences can be further divided into *discovery* – the process of creating new theories or laws – and *justification* – the testing of the validity of these new theories or laws. Design science, on the other hand, attempts to create artificial artifacts that serve human purposes. Design science can be divided into two activities: *build* and *evaluate*. The first activity concerns the building of an artifact for a specific purpose, and the second deals with evaluating the success of the artifact according to different devised measures.

Research outputs are divided into four categories. *Constructs* form the specialized vocabulary of the domain, which is used to describe research problems. Constructs can be formalized (as in semantic data modeling) or informal (as in cooperative work). A *model* is composed of constructs, and provides problem and solution statements. A *method* is an algorithm used to perform a task, and is based on a set of underlying constructs, and a model of the solution. An *instantiation* is a realized artifact, and demonstrates the feasibility and effectiveness of the underlying models and methods. Especially in IT research, the instantiation might actually precede the underlying constructs,

---

[2] Constructive research is here taken as the approach as defined by Kasanen et al. (1993), not to be confused with the constructive view of learning as used in education.

[3] Design science in IT can be defined as research that "creates and evaluates IT artefacts intended to solve identified organizational problems" (Hevner et al. 2004) and the constructive approach as "problem solving through the construction of models, diagrams, plans, organizations, etc." (Kasanen et al. 1993). Taking the view that artefacts can consist of constructs, models, methods, and instantiations (March and Smith 1995), in an IT setting the constructive methodology and design science can be viewed as largely the same. This will be the view adopted from here on in this thesis.

models, and methods, as demand for a particular artifact might cause its conception using just intuition and experience. The constructs, models, and methods can then be studied, and better instantiations developed. (March and Smith 1995)

		Research Activities			
		Design Science		Natural Science	
		Build	Evaluate	Theorize	Justify
Research Outputs	Constructs				
	Model				
	Method				
	Instantiation				

Figure 2.1. March and Smith's Research Framework (March and Smith 1995)

According to Nunamaker et al. (1991), much information systems research follows a basic *research life cycle* of the form *concept – development – impact*. This implies that basic research is followed by experimental implementation, and finally, by research into user acceptance and productivity. The constructive approach can be seen as an integral part of this chain, in the development stage. Following the constructive approach, a concept introduced through previous research can be applied to solve a specific problem, usually through the development of a physical artifact. Although the actual functioning of the artifact must be evaluated and proven, it will be the topic of future research to determine the impact of the innovation.

2.1.2. Research in this dissertation

The research in this thesis positions itself in the *design science* column of the March and Smith (1995) framework (Figure 2.1), in both the *build* and *evaluate* columns. Specifically, a *model* for competitive financial benchmarking using self-organizing maps is built and evaluated. In Järvinen's tree-like taxonomy of research approaches, this research is located in the *innovation building and evaluating approaches* branch (Figure 2.2). This thesis will thus use the *constructive research approach*.

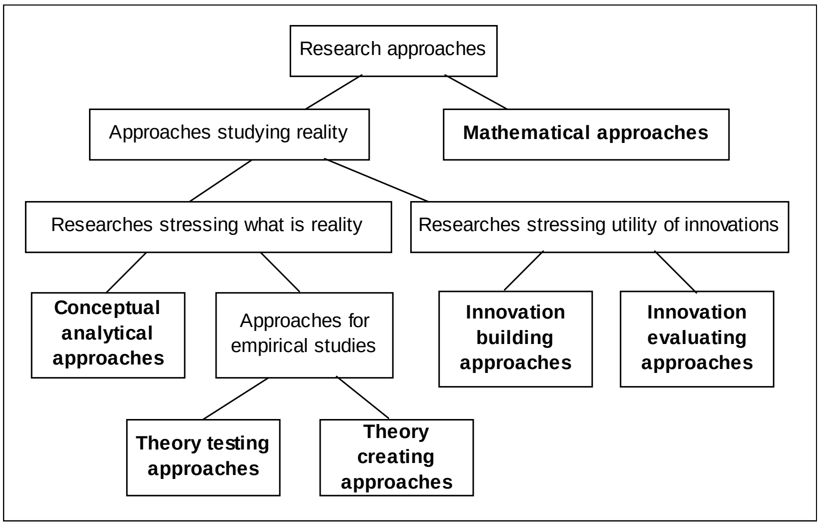


Figure 2.2. Järvinen and Järvinen's Taxonomy of Research Methods (Järvinen 2001, p.10).

March and Smith regard the building and evaluative activities of design science as separate activities, claiming "The research contribution lies in the novelty of the artifact and in the persuasiveness of the claims that it is effective. Actual performance evaluation is not required at this stage" (March and Smith 1995). The constructive approach, however, requires that the value and the actual working of the construct be shown as well (Kasanen et al. 1993; Järvinen 2001).

March and Smith (1995) state that research in the evaluate activity requires the development of suitable metrics. The authors propose five evaluation criteria for evaluating models. These are *their fidelity with real world phenomena*, *completeness*, *level of detail*, *robustness*, and *internal consistency*. To these Järvinen (2001, p.111) adds: *form and content*, *richness of knowledge*, and *user experiences*. Lincoln and Guba (2000, p.170) mention that the primary goodness or quality criteria according to the constructive approach are *trustworthiness* and *authenticity*. In this thesis, the self-organizing map will be evaluated as a financial benchmarking tool based on its fidelity with real phenomena, i.e. how well the results reflect reality, and richness of knowledge, i.e. how much information about reality is provided by the model. A number of subject matter experts will assess the model according to these criteria through a face validation procedure, in order to determine the model's trustworthiness and authenticity.

This thesis is based on a growing body of research on self-organizing maps, initiated by Kohonen in 1972 (1972; 1997; 2001). Research in the same area (Back et al. 1995; 1996; 1998a; 1998b) has also motivated the research in this thesis. Also, technological advances have made it possible to use self-organizing maps on personal desktop computers. Likewise, an increasing base of financial information has created a demand for a new data-mining tool, motivating the evaluation of the functioning of this model. (Järvinen 2001)

Specifically, this thesis utilizes the innovation building and evaluating approaches. Innovation-evaluating approaches seek to evaluate the usefulness of a tool for a specific task (Järvinen 2001). In this case, the purpose is to evaluate the self-organizing map as a financial benchmarking tool, by creating a benchmarking model for evaluation by a number of subject matter experts, i.e. potential end-users of the application.

According to Kasanen et al. (1993), the constructive research process is the following:

1. Find a practically relevant problem, which also has research potential.
2. Obtain a general and comprehensive understanding of the topic.
3. Innovate, i.e. construct a solution idea.
4. Demonstrate that the solution works.
5. Show the theoretical connections and the research contribution of the solution concept.
6. Examine the scope of applicability of the solution.

In this case, the relevant problem is the lack of a feasible tool for analysis of large amounts of publicly available financial information (1). The access to financial information has increased greatly during the past few years, mainly due to the Internet. One possible tool for this purpose is the self-organizing map, which has been applied to a wide range of problems, and in some cases, to specifically financial problems (see Section 1.3). A study of current methods for financial benchmarking in Finnish publicly-noted companies has shown that very few advanced methods are currently used, and managers' satisfaction with their current methods leaves room for improvement (2). The problem to be studied, therefore, is to build and evaluate a model for financial benchmarking in the international pulp and paper industry using the self-organizing map. An understanding of the topic will be obtained by studying existing literature on the subject, as well as by studying the three basic concepts of the study: *benchmarking*, *knowledge discovering in databases* (specifically: *data mining using neural networks*), and *financial analysis*. A solution that combines the three concepts above to provide a potential element for a decision support system for financial managers, analysts, and stakeholders will be presented (3). The functioning of the solution will be demonstrated through a face evaluation of the model by a number of subject matter experts (4). Finally, the conclusions based on this experiment will be presented, and some thoughts on the future of self-organizing maps in financial analysis will be described (5 and 6). The constructive research process as applied in this thesis is presented in Figure 2.3.

1. Find a relevant research problem.	Chapters 1 and 3, *papers* 1, 2, and 5.
2. Obtain a comprehensive understanding of the topic.	Chapters 3, 4, 5, and 6.
3. Innovate.	Chapter 8, papers 1, 2, and 3.
4. Demonstrate that the solution works.	Chapters 8 and 9, papers 1, 2, 3, 4, and 5.
5. Show the theoretical connections.	Chapters 2, 3, 4, 5, 6, and 8.
6. Examine the scope and applicability of the solution.	Chapter 10, papers 1, 2, 3, and 4.

Figure 2.3. The Constructive Research Process in this Thesis (Kasanen et al. 1993).

Hevner et al. (2004) define seven guidelines for design science (constructive research) in IS research. The construct must be innovative and purposeful (i), applied for a specified problem domain (ii). Evaluation is crucial (iii) and the construct must solve a heretofore unsolved problem, or solve a known one in a more effective or efficient manner (iv), thereby differentiating it from the practise of design. The artefact must be rigorously defined, formally represented, coherent, and internally consistent (v). Problem solving should use available means to reach desired ends while satisfying laws existing in the environment (vi). Finally, the results must be effectively communicated to an academic audience as well as practitioners and managers (vii).

Based on these guidelines, the research in this thesis is analyzed as follows. The construct in question is a new financial benchmarking model based on the self-organizing neural network (i) intended to assist managers in the international pulp and paper industry performing competitor financial benchmarking (ii). Evaluation in the form of an expert survey and face validation is used to evaluate the construct (iii) and to demonstrate that the construct offers a new and better solution to an existing problem (iv). Rigor will be derived from the effective use of the existing knowledge base (i.e. prior research and valid research approaches), both in the construction and the evaluation of the construct (v). Problem solving has been performed in an iterative pattern, by searching for the best possible means to solve the problem and continuously testing / generating solutions (as witnessed in the iterative, developmental publications), and finally evaluating the construct from a business perspective (vi). Finally, knowledge of the results have been presented to both academic audiences (in the form of this dissertation and its publications, and through conference participation (e.g. Eklund et al. 2002)), as well as to managers and practitioners in industry publications (Eklund et al. 2001) and company-financed research projects (vii).

This research, therefore, conforms to the guidelines proposed by Hevner et al. (2004), and can be classified as design science.

2.2 Verification and Validation

A central component of the constructive research approach is the evaluation of the created constructs (Kasanen et al. 1993; March and Smith 1995; Järvinen 2001). Verification and validation are two aspects of evaluation. According to the definition used by for example the US Department of Defense, verification is "[t]he process of determining that a model implementation accurately represents the developer's conceptual description and specifications", while validation is "[t]he process of determining the degree to which a model is an accurate representation of the real-world from the perspective of the intended use of the model" (Page et al. 1997). Verification can be seen to constitute guideline four, and validation guideline five, in Hevner et al.'s (2004) guidelines for design science (see Section 2.1.2).

In this case, verification is not a formally defined process, but is more an implicit part of the research approach. This is because the approach requires strong documentation, and motivations from extensive literature surveys, i.e. a rigorous research process. In addition, the results of the final construct are compared to existing domain knowledge, in the form of industry publications and the textual parts of the annual reports used.

The model will be evaluated through an expert survey. The survey will take a face validation approach, i.e. subject matter experts (SMEs) are surveyed concerning their opinions of a model demonstrated to them. This is necessary as the model that is being evaluated is based on the use of several different software programs. As there is no existing prototype of a user-interface for the model, it would be impossible to allow the users themselves to experiment with the tool at this stage.

Surveys are quantitative techniques used to measure some specific aspects of a study population, usually through structured, predefined questions administered to a defined sample of the population (Galliers 1992; Pinsonneault and Kraemer 1993). The validation of the model in this case includes two surveys. The first of these, the expert survey of current methods (Section 3.2) includes elements of both *exploratory* and *descriptive* survey research. The purpose of exploratory survey research is to become more familiar with a topic and to test preliminary concepts about it, whereas in descriptive survey research, the objective is to determine what situations, events, attitudes, or opinions are occurring in a population (Pinsonneault and Kraemer 1993). The first survey is exploratory in that it aims to explore what methods are currently being used for financial benchmarking, and descriptive in that the users are queried as to their attitudes towards these methods. The objective is to provide a domain basis for the second survey, and to assess the novelty of the benchmarking model. The second survey (Chapter 9) is clearly an example of a descriptive survey, as it seeks to evaluate managers' attitudes towards a specific model.

2.3 Evaluation of Information Systems

There has been a very lively discussion concerning the proper way to evaluate the success of information systems in the literature. For a long time, researchers have sought to find the elusive dependent variable of information systems success (DeLone and McLean 1992). For example, DeLone and McLean identified over 100 different measures used to evaluate IS success in the 180 studies that they collected. The authors identified six different categories of measures of IS success that had been used in the studies: *system quality, information quality, use, user satisfaction, individual impact,* and *organizational impact*. DeLone and McLean note that these measures can be arranged into a success construct as illustrated in Figure 2.4. The model shows the interdependent nature of IS success evaluation. The authors also note that "no single measure is intrinsically better than another; so the choice of a success variable is often a function of the objective of the study, the organizational context, [etc.]" (DeLone and McLean 1992). DeLone and McLean's model is easily one of the most influential models, as it has been cited in over 285 refereed journal papers since mid 2002 (DeLone and McLean 2003). In their update of the influential 1992 paper, the authors listed a number of studies that had attempted to verify the casual relationships between the different measures of success in the model, but again noted that few studies used multiple measures of success as suggested in their original paper (DeLone and McLean 2003).

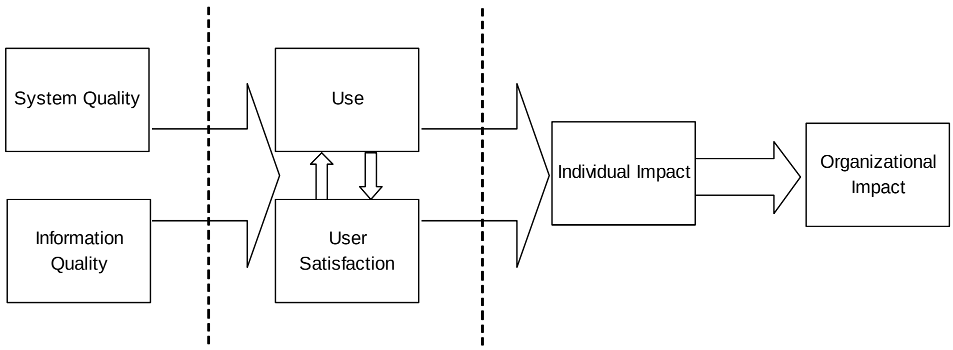


Figure 2.4. DeLone and McLean's (1992) IS success model.

In recent years, the discussion has primarily centered around three different measures of success: *cost-benefit analysis, system usage,* and *end-user satisfaction* (Au et al. 2002).

Of these, cost-benefit analysis, centering on measures such as decreased costs, improved processing times, etc., is seemingly the most objective approach. However, these effects are also often difficult to measure, as there can be many factors that influence them. For example, monetary estimates of the effects of an IS can be notoriously difficult to make, especially when the effects are intangible, such as improved customer support (Murphy and Simon 2002). Also, in many cases it is difficult to prove that a particular benefit or cost can be attributed solely to a new information system (Au et al. 2002). Finally, the objectivity of these measures has also been questioned, as the underlying estimates are themselves based on experts' subjective predictions, for example, estimated payback and durations of implementation (Saarinen 1996). In many cases these measures are, therefore, simply not feasible.

Usage-based evaluation adopts the view a system is successful if it has a high degree of use (Gelderman 1998). Indeed, Downing (1999) found that usage behavior analysis yielded similar results as the validated user satisfaction model by Doll and Torkzadeh (1988). However, this view has also been criticized in the literature, as IS use might not be entirely voluntary. Users may dislike the system, yet might still be forced to use it daily. Therefore, it is argued that system usage as a measure of IS success should only be used when the use of the system is entirely voluntary (Au et al. 2002). DeLone and McLean (2003), however, argue that no system use is entirely mandatory, and that, for example, declining usage can be a clear indication of a system not yielding the anticipated benefits. Therefore, while a too simple measuring of time of use is not feasible, system use can in certain circumstances be a feasible measure of system success.

If taking the view that system use is a valid measure of IS success, Davis' (1989) *Technology Acceptance Model* (TAM) can be used to predict the level of use of a particular technology. The TAM model (Figure 2.5), further developed and validated in Davis et al. (1989), is based on two fundamental concepts; *perceived usefulness* and *perceived ease of use*. The user must perceive that the technology will help her perform her job better than before, and that the effort of using the technology will not outweigh the benefits (Davis 1989). This will determine the user acceptance of the technology. The 12-item TAM model has been widely accepted and validated in the MIS literature (Doll et al. 1998). A particular advantage of the model is that it can be used to predict use, and is therefore well suited for evaluating prototypes used in face validation settings.

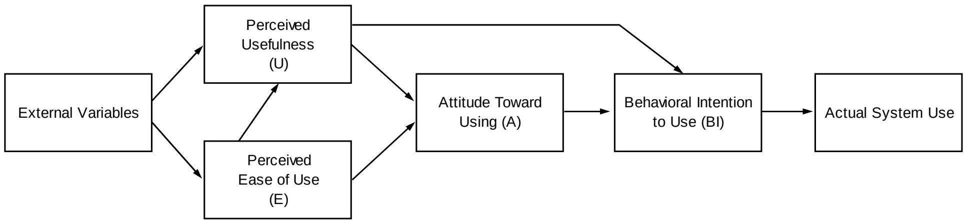


Figure 2.5. Davis' (1989) Technology Acceptance Model (TAM) (Davis et al. 1989).

User information satisfaction (UIS)[4] has emerged as the most widely used standard for measuring the success of information systems (Melone 1990; DeLone and McLean 1992; Au et al. 2002). Baroudi et al. (1986) present evidence that higher user satisfaction does indeed lead to greater system usage. For the reasons presented earlier, many researchers have claimed that user satisfaction is a better measure than cost-benefit analysis and usage-based evaluation (Au et al. 2002). User satisfaction stresses the importance of perceived system and information quality rather than technical quality (Ives et al. 1983), making it a suitable measure in most cases. Many different models for evaluating user satisfaction have been proposed in the literature (e.g. Bailay and Pearson 1983; Ives et al. 1983; Adelman et al. 1985; Baroudi and Orlikowski 1988; Doll and Torkzadeh 1988; DeLone and McLean 1992; Doll et al. 1995; Etezadi-Amoli and Farhoomand 1996; Saarinen 1996; Seddon 1997; Gelderman 1998; Goodhue et al. 2000; Au et al. 2002; Rai et al. 2002; Muylle et al. 2004), each emphasizing different aspects of satisfaction, in addition to other aspects of success.

However, user satisfaction has also been criticized in the literature. For example, Goodhue (1995) claims that better IS performance does not necessarily lead to higher user satisfaction. This is problematic as user satisfaction is generally seen as a supplementary measure of IS success. On the other hand, many researchers claim that even a "good" system perceived as poor by its users is indeed poor (Ives et al. 1983). Also, the validity of a number of the validated models has been questioned (Galletta and Lederer 1989; Melone 1990; Saarinen 1996; Au et al. 2002). For example, the UIS construct (Bailay and Pearson 1983) does not directly measure the impact of an IS on the individual or the organization (Melone 1990; DeLone and McLean 1992; Saarinen 1996). Generally, however, there is a consensus in the field of IS that user satisfaction is more often a feasible measure than the other alternatives (e.g. Saarinen 1996; Igbaria and Tan 1997; Gelderman 1998; Au et al. 2002).

[4] User satisfaction (US), end-user satisfaction (EUS), end-user computing satisfaction (EUCS), end-user information satisfaction (EUIS) are, for our purposes, similar concepts in the literature.

As was previously mentioned, several different models for assessing user satisfaction have been proposed in the literature. One commonly used model is the end-user computing satisfaction (EUCS) model, proposed and validated by Doll and Torkzadeh (1988). The authors used factor analysis to isolate the five most important factors in end-user computing satisfaction, and thereby validate their 12-item satisfaction questionnaire. These factors are *content, accuracy, format*, *ease of use*, and *timeliness*. Doll and Torkzadeh's model has been widely validated and used, often in adapted form (for example, Seddon and Yip 1992; Doll et al. 1994; Igbaria and Tan 1997; Downing 1999; Gordon and Geiger 1999; Chen et al. 2000b; Wu et al. 2001; Somers et al. 2003; Ong and Lai 2004).

In this evaluation, the only feasible measure of success is user satisfaction. The Doll and Torkzadeh (1988) model (Figure 2.6) has been used as the basis for the construction of the questionnaires. The primary motivation for using this framework in this case is that it focuses more on information quality questions (Li 1997; Au et al. 2002) than the other validated models available, many of which try to assess other aspects of performance as well. In a number of studies, information quality has been proven to have a strong effect on system use and net benefits, and DeLone and McLean (2003) encourage the use of information quality measures in any success measurement construct. As this validation was a face validation, and the users would not be using the systems themselves, models that emphasize information quality were preferable over models that emphasize technical properties or organizational impact, as these are extremely difficult to predict. Also, the EUCS tool is more suitable for the evaluation of a specific tool than, for example, Bailey and Pearson's or Ives et al's models, as these include questions concerning EDP staff and services and user involvement questions (Doll and Torkzadeh 1988), which are not applicable in this validation process. Studying DeLone and McLean's model in Figure 2.4, the current research lays in the temporal dimension somewhere in between information quality and user satisfaction, but any further implications of the tool would be impossible to assess in this setup. Davis' TAM model could also have been applicable but not sufficient, as the emphasis was on specific information quality-related questions. The TAM model is quite general, and would probably not have been able to point out specific weaknesses in the SOM model itself. The TAM model would have been very useful in a domain specific prototype-setting, i.e. a prototype of the model would have been used by representatives of a specific industry, which was not possible in this case. However, the concepts of perceived usefulness and perceived ease of use are central to this research, and have therefore also been included in this questionnaire. These reasons made the Doll and Torkzadeh model preferable in this case.

However, the Doll and Torkzadeh model is not as such entirely applicable to this case, and has therefore been adapted to measure the aspects directly relevant in this case. The authors' definition of end-user satisfaction is "the affective attitude towards a specific computer application by someone who interacts with the application directly”. As such, the users are not yet direct users, so this required some changes in the instrument. While the model still primarily builds upon the five-factor model from Doll and Torkzadeh (1988), it has been supplemented with questions from the other satisfaction studies and models, such as Bailay and Pearson (1983), Seddon and Yip (1992), and Alter (2002, p.163).

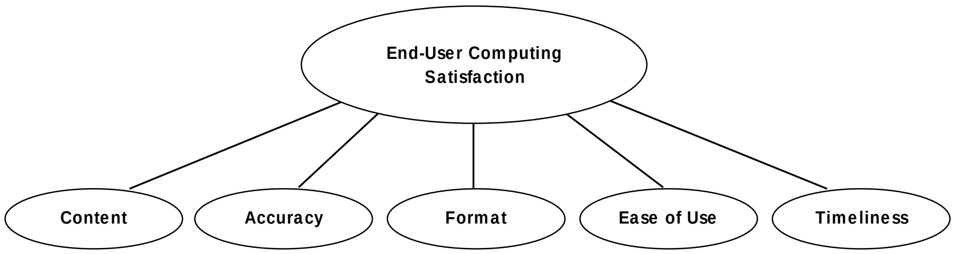


Figure 2.6. Doll and Torkzadeh’s (1988) model of end-user computing satisfaction.

3 KEY CONCEPTS AND STATE OF THE ART IN FINNISH COMPANIES

In this chapter, a number of key concepts for this thesis will be briefly presented. These concepts are central to thesis in that they represent a number of different possible approaches to dealing with financial data in companies.

The second part of the chapter describes the results of a survey conducted to determine the current state of the art in financial benchmarking in Finnish, publicly-noted companies, as well as managers' satisfaction with these methods. Finally, information overload and the complexity of the competitive environment will also be assessed.

3.1 Key Concepts and Competing Technologies

3.1.1. Knowledge Discovering in Databases and Data Mining

Knowledge discovering in databases (KDD) can be defined as "the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data" (Fayyad et al. 1996). There is some controversy concerning the definitions of the terms knowledge discovering in databases (KDD) and data mining. In much of the literature on data mining, the terms are held as largely synonymous. However, some authors, for example, Fayyad et al. (1996) and Adriaans and Zantinge (1996), define knowledge discovering in databases (KDD) as the entire process of extraction of knowledge from the data, and data mining as only the actual use of intelligent tools during the discovery stage of the process. This is also the definition of the International Conference on KDD and data mining, first held in 1995. Since the later definition allows for a better distinction of the actual data-mining step it will be used in this thesis.

The KDD process can be illustrated as a five-step process (Figure 3.1):

1. Data selection
2. Data preprocessing
3. Data mining
4. Reviewing output
5. Interpreting results

During the data selection stage, the data sources to be used are selected. The data can be collected from a number of sources, for example, databases, data warehouses, newswire feeds, flat files, etc.

During the following stage, the data are preprocessed, which implies collecting and cleaning the data. Cleaning is done using filtering or error detection programs to remove incomplete, incorrect, or duplicate data. The data might also be enriched with data from external sources. Finally the data are coded.

In the third step, the actual data mining application is run. An analyst then analyzes the results (step 4), and might refine or revise the original query. Once the analyst is pleased with the output, the results are interpreted (step 5), and actions are taken based on the output. (Adriaans and Zantinge 1996; Fayyad et al. 1996)

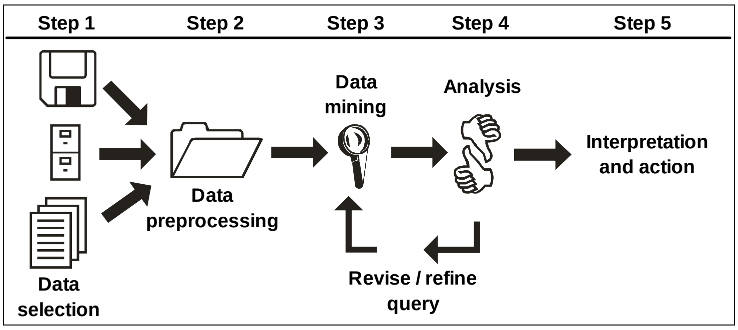


Figure 3.1. The KDD process. (Adapted from Adriaans and Zantinge 1996; Fayyad et al. 1996)

3.1.2. Data Mining

Data mining can be defined as "a step in the KDD process consisting of applying data analysis and discovery algorithms that, under acceptable computational efficiency limitations, produce a particular enumeration of patterns over the data" (Fayyad et al. 1996). The principle difference between data mining applications and analytical tools is that data mining allows us to find optimal clusterings, or regularities in a database (Adriaans and Zantinge 1996). Data mining is used to find information that cannot be discovered using standard database queries such as "How many companies have a return on equity > x and a current ratio > y?" An example of a question more suitable for data mining applications is "which companies exhibit the best overall financial performance compared to other companies?"

Most analytical tools support a verification-based approach. This implies that the user formulates a hypothesis about data interrelationships, and then uses the tool to verify or refute the hypothesis. This approach requires that the analyst is able to intuitively pose the appropriate questions, and is able to correctly formulate these questions as potentially very complex queries (Moxon 1996). For example, SQL, a popular and common analytical tool, is only a simple query language; it only allows us to find data under constraints that we already know (Adriaans and Zantinge 1996, p.7).

Data mining, as opposed to analytical tools, uses discovery-based approaches to find key relationships in the data (Moxon 1996; Fayyad et al. 1996). This implies using pattern-matching and other similar algorithms to find hidden relationships, which are invisible to standard database queries. Discovery-based approaches can be further divided into prediction and description, based on the overall goal (Fayyad et al. 1996).

Information technology is essential to data mining. When discussing the task of knowledge generation Spiegler (2003) states that technology is the prerequisite and means in this task. Spiegler also claims that "if data become information when they add value, then information becomes knowledge when it adds insight, abstraction, and better understanding".

There are several methods, or approaches, for data mining. The choice of method used depends upon the type of information to be extracted (Fayyad et al. 1996). Different techniques include *regression, dependency modelling, sequence analysis, classification, clustering,* etc. Traditional back-propagation neural networks are classification tools, i.e. they use a set of pre-classified examples to develop a model that can classify the population of records at large. Self-organizing maps are clustering tools, i.e. they create a map on which similar groups of records are grouped close to each other, and natural boundaries between these groups are defined. Self-organizing maps are presented in detail in Chapter 6.

3.1.3. Data Envelopment Analysis

Data envelopment analysis (DEA) is an efficiency comparison method based on linear programming that was proposed by Charnes et al. (1978) and further developed by Banker et al. (1984). DEA is concerned with comparing the relative efficiency of a number of decision making unit (DMUs) with similar goals and objectives, in order to find a single overall best performing unit, and thereby, to identify best observed practices (Athanassopoulos and Ballantine 1995). DEA is particularly useful in cases where it is difficult to assign prices to many of the outputs, such as in hospital and government efficiency evaluation (Steering Committee for the Review of Commonwealth/State Service Provision 1997).

A central concept in DEA is the *efficient frontier*. Figure 3.2 shows an example (Adapted from Anderson 1996) of what an efficient frontier is. The figure shows a single (equal) input / two output comparison, with DMUs producing the two outputs in differing amounts. In the figure, A, B, and C are efficient DMUs, as they define the maximum outputs among the illustrated DMUs. They thus form an efficient frontier, and any DMUs on this frontier are considered efficient (have an efficiency of 1, or 100%). D, E, F, and G are thus inefficient, as they could theoretically increase their output. For example, G could increase its output to G' (3.5, 7) and thereby reach the efficient frontier. The term envelopment refers to that inefficient DMUs are located inside an area (shaded area in Figure 3.2) enveloped by the efficient DMUs (Doyle 1995). The figure also incorporates the improvement suggested by Banker et al. (1984), i.e. that of scale effects. Beyond a certain input, the output does not linearly increase.

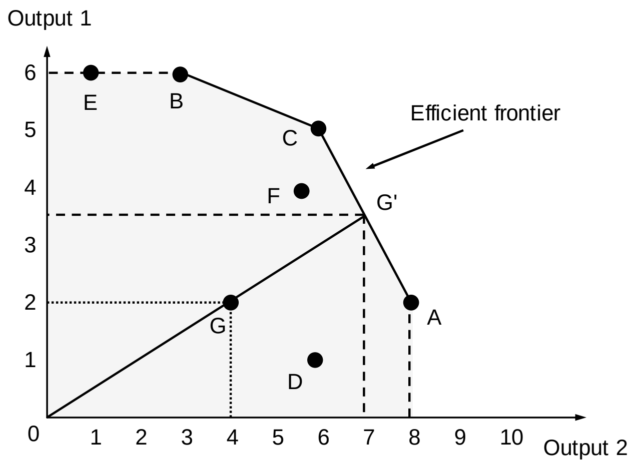


Figure 3.2. A sample DEA efficient frontier (Adapted from Anderson 1996).

The example above is a very simple example. In reality, DEA is applied to much more complex cases, using multiple inputs and outputs.

DEA is based on the concept of relative efficiency, which can be defined as the ratio of the weighted sum of outputs to the weighted sum of inputs (Emrouznejad 2003). The problem is that a common set of weights is often impossible to assign. For example, one DMU might prioritize different outputs than another. Therefore, in DEA each DMU sets its own weights in a way that maximizes its own efficiency score (Emrouznejad 2003). This process is referred to as self-evaluation and cross-evaluation; each DMU tries to make itself as attractive as possible compared to other DMUs (Doyle 1995). This is solved as a series of linear programming problems, which are combined to give an overall efficiency score for each DMU. The result is an efficient frontier as in Figure 3.2, against which the efficiencies of the DMUs are compared, showing their relative efficiencies. This efficiency score shows if the unit is efficient (the score is = 1, or 100%), or if it is inefficient (the score is < 1), how far it is from being efficient. The weighting can also be based, for example, on decision maker's preference structures (Zhu 1996; Yun et al. 2004), the stability of the efficiency of a DMU (Charnes et al. 1996), or a type of “average” weighting inferred from the data themselves (Doyle 1995).

There are a number of benefits and drawbacks of using DEA. DEA is capable of handling complex multiple input, multiple output problems, and does not require the user to specify the weights (Anderson 1996). However, care is required in analysis of the results, as the self-evaluation function will allow units that are producing only one output in great amounts to be listed as particularly efficient, even though this might not be feasible in reality (Emrouznejad 2003). Also, as the output is either efficient or inefficient, the characteristics of a particular DMU might be difficult to judge.

DEA is an extreme point method, measuring the performance of DMUs against the best performers, not against averages as is typical for statistical methods (Anderson 1996). This makes DEA a good tool for the performance benchmarking process, as by definition, we are comparing against the best performing units. On the other hand, extreme point methods are very sensitive to outliers and mistakes, something that is common in financial data.

Dealing with negative values also requires additional processing, as DEA does not work with negative values (Feroz et al. 2003). This makes working with profit/loss figures difficult. Finally, DEA can be quite computationally demanding, as a separate linear program for each DMU is created. This would make working with the amounts of data included in this experiment quite cumbersome.

DEA has become a very popular topic in the scientific literature. From 1978 to 2001, 3,203 articles were written concerning DEA, 1,259 of these in journals (Tavares 2002). It has been applied in a wide range of applications, for example, bankruptcy prediction (Cielen et al. 2004), resource allocation and efficiency (Chen 1997; Athanassopoulos and Gounaris 2001), financial statement analysis (Smith 1990), and technology selection (Baker and Talluri 1997). In particular, DEA has been applied in a variety of benchmarking applications, such as selecting performance benchmarks (Post and Spronk 1999; González and Álvarez 2001; Rickards 2003) and performance evaluation in various industries (Athanassopoulos and Ballantine 1995; Homburg 2001; Feroz et al. 2003), just to name a few.

3.1.4. Balanced Scorecards

The balanced scorecard is not in itself a financial benchmarking method, but is actually a strategic management tool. The balanced scorecard was originally proposed by Kaplan and Norton (1992; 1993), and was based on the idea that managers should not have to choose between financial and operative measures of success. The authors suggested that the business should be studied from four important perspectives: the *customer, financial, internal business*, and *innovation and learning perspectives* (Kaplan and Norton 1992). The scorecard thus emphasizes a balance between internal and external measures. A core thought is that each perspective should be assessed using a very restricted number of measures. The total number of measures should not exceed 15-20 (Kaplan and Norton 1993).

Rockwater, Apple, and Advanced Micro Devices (AMD) are examples of early adopters of balanced scorecards (Kaplan and Norton 1993).

Although balanced scorecards are not specifically financial benchmarking methods, they are interesting because they represent a holistic view of benchmarking, including both internal and external measures. This is why it is important to assess their degree of use in Finnish companies.

3.1.5. Other techniques

There are a number of other possible techniques for financial competitor benchmarking, primarily from the field of statistics. Statistical techniques, such as clustering techniques, time-series analysis, regression analysis, etc., are well-known techniques, and will therefore not be further discussed in this thesis. The primary advantage of using statistical tools is that they have a strong foundation in the literature, and are commonly accepted as functioning under certain assumptions. Some limitations include their lack of advanced visualization properties and their dependence upon parametric statistics, under assumptions that are difficult to satisfy in real applications.

There are also a large number of other data-mining tools that could be used, but they also often lack the visualization capabilities of the SOM.

3.2 Survey of experts

In order to determine the feasibility of using the SOM for financial benchmarking, it was important to first determine the current state of the art in business. It was also important to determine the managers' satisfaction with their current methods, in order to determine if there is a demand for more advanced tools. In order to answer these questions, an expert survey of Finnish HEX-noted companies was conducted. The online survey was conducted between October and November 2003. The survey is presented in detail in Eklund et al. (2004, Paper 1).

3.2.1. Research design

The questionnaire was constructed based on the Doll and Torkzadeh (1988) model as discussed in Section 2.3, with some modifications. The questionnaire was split into five parts: *demographics and degree of IT familiarity, current sources and methods for data analysis, importance of factors of information, satisfaction with current methods,* and *information overload and complexity*. The respondents were also asked if they were interested in taking part in a second phase, the demo and evaluation of the benchmarking model (Chapter 9). The questionnaire was administered in English, as most of the companies were multinational in their operations. Most of the questions were based on 5 point Likert scales (1 = strongly disagree, 2 = somewhat disagree, 3 = neutral, 4 = somewhat agree, 5 = strongly agree). There were also a number of other attribute scales, such as very important – very unimportant, very satisfied – very dissatisfied, etc., also on 5 point scales. Finally, there were a number of open questions. The survey was first tested on academic colleagues and three independent respondents from industry, after which a number of modifications were made to the original survey.

3.2.2. Sample and administration

Finnish publicly-noted companies, restricted to those noted on the main HEX-index[5], were used as the population of the study. So-called, I-, NM-, or pre-list companies were left out, as these are often small, non-international companies. The total population included 103 companies representing a large variety of different industries. Targeted managers were contacted personally before sending the questionnaire. At this stage, ten companies elected not to participate, motivating their decisions by time constraints or lack of applicability to their own companies. Thus, a population of 93 companies remained. Within the deadline set, and after two e-mail reminders, 38 answers were received, representing 40.86% of the sent questionnaires (36.89% of the original total population).

3.2.3. Results

Although background information was optional, 73.68% of the respondents provided this information. The background information showed that the managers generally had a high level of education (78.57% had an M.Sc. or higher), and 76.67% were directly involved in business intelligence, strategic development, or corporate finance. The managers were also fairly experienced, often having worked in the same company for a long time. However, the variance was high, indicating many experienced managers, but also a number of junior managers. The managers were experienced users of basic IT applications such as

[5] Helsinki Stock Exchange

word processing, spreadsheets, e-mail, the Internet, and electronic calendars. These tools were also often used daily. However, the managers reported lower experience with databases, and very low experience with decision support systems. In fact, the decision support systems represented the largest beginner group of the study. This is relevant as these applications have relevance to the SOM model presented in this thesis, showing that the managers have had little experience of similar applications.

Familiarity with IT tools:	N	Mean	SD	Med	Expert – Average – Beginner 5	4	3	2	1
Word processing	38	3.84	0.55	4	7.89%	**68.42%**	23.68%	0.00%	0.00%
Spreadsheets	38	3.87	0.81	4	18.42%	**57.89%**	15.79%	7.89%	0.00%
E-mail	37	4.05	0.62	4	21.62%	**62.16%**	16.22%	0.00%	0.00%
Calendars	38	3.39	1.10	4	10.53%	**47.37%**	21.05%	13.16%	7.89%
Databases	38	3.13	0.99	3	7.89%	26.32%	**42.11%**	18.42%	5.26%
Internet	38	3.95	0.61	4	15.79%	**63.16%**	21.05%	0.00%	0.00%
Decision support systems[6]	35	2.74	1.24	3	5.71%	25.71%	**28.57%**	17.14%	22.86%

Table 3.1. Familiarity with common IT tools.

The most important objectives of the study were to assess which sources were commonly used to obtain financial information, what methods were used to analyze these data, and how satisfied managers were with the methods in current. The survey found that the most common sources of financial information were newspapers. This finding was interesting, as at least one manager questioned the reliability of this source in the open answers. The Internet was the second most commonly used source, and industry publications and informal channels were also commonly used. Financial reports were most commonly used on a quarterly basis. Interestingly, external reports were very infrequently used, and the only source that received any "never use"-answers. The conclusion is therefore that most of the surveyed companies seem to do competitor analysis "in-house", instead of relying on external services. This increases the relevancy of this study in this evaluation setting.

The tools used by the managers remained very basic. The most commonly used tools were spreadsheets. Nearly all managers used spreadsheets at least a few times per quarter. Statistical tools, balanced scorecards, and weighted averages of financial ratios were considerably less commonly used, even on a quarterly basis. Data envelopment analysis and neural networks were very uncommonly used tools, with 91.89% of managers responding that they rarely or never use these

[6] Decision support systems were in this case primarily defined as information systems that present data from a database in a form suitable for decision making.

tools. In other words, very few advanced, multiple-ratio methods are used for financial competitor benchmarking by the managers. Managers primarily rely upon spreadsheet analysis of financial ratios, i.e. the one-ratio-at-a-time approach.

Table 3.2 shows the managers' satisfaction with current financial benchmarking methods on a scale of 1-5, with 1 being very dissatisfied and 5 being very satisfied.

					Very satisfied – Neutral – Very dissatisfied				
Variable	N	Mean	SD	Med	5	4	3	2	1
Content	38	3.34	0.75	3.5	0.00%	**50.00%**	34.21%	15.79%	0.00%
Accuracy	38	3.29	0.73	3	0.00%	**44.74%**	39.47%	15.79%	0.00%
Format	38	3.13	0.74	3	0.00%	34.21%	**44.74%**	21.05%	0.00%
Ease of Use	38	3.18	0.87	3	0.00%	**42.11%**	39.47%	13.16%	5.26%
Timeliness	38	3.34	0.91	4	2.63%	**52.63%**	23.68%	18.42%	2.63%
Overall	37	3.24	0.80	3	0.00%	**45.95%**	32.43%	21.62%	0.00%

Table 3.2. Manager's satisfaction with current methods.

The managers, in absence of experience of the SOM in benchmarking, appear to be relatively satisfied with most factors of current methods for financial benchmarking. However, there does appear to be a central tendency in the answers, and managers do not appear to be overly convinced in their answers. For example, the only factor to get any strongly agree-answers was timeliness, and it only received 2.63%. On the other hand, the variance for timeliness is very high, so it appears that there are great differences among the companies concerning the timeliness of their methods. There are also a fairly large number of managers dissatisfied with some aspects of current methods. In general, the lowest satisfaction was reported with two factors: format and ease of use. This is important, as arguably the main contribution of the SOM model would be specifically in these two factors.

The managers were also surveyed regarding information overload and the complexity of the competitive environment. The level of frustration with daily information was split fairly evenly among the managers. There were a fairly large number of managers who indicated frustration. This provides support for the application of new tools for financial benchmarking.

					Constantly frustrated – Neither – Never frustrated				
Frustration with daily information	N	Mean	SD	Med	5	4	3	2	1
Frustration	38	2.87	1.04	3	2.63%	31.58%	23.68%	**34.21%**	7.89%

Table 3.3. Frustration with daily information.

Competitive environment	N	Mean	SD	Med	Very complex – Neither – Very uncomplex 5	4	3	2	1
Complexity	38	3.87	0.93	4	23.68%	**52.63%**	10.53%	13.16%	0.00%

Table 3.4. Complexity of the competitive environment.

Managers indicated a high degree of complexity in the competitive environment. Plumlee (2003) studied the effect of information complexity on analysts. Her results indicated that analysts assimilate less complex information to a greater extent than they assimilate more complex information. Plumlee studied taxation documentation, and as such, her results are not comparable to this case. However, if increasing dimensionality is viewed as increasing complexity, then dimensionality reducing methods, such as clustering methods and specifically visual clustering methods, could be viewed as complexity reducing tools.

In a classic study, Miller (1956) studied the number of pieces of information that humans can simultaneously store and process. His conclusion was the limit is seven plus/minus two pieces of information at a time. He also concluded that by processing information in "chunks", more information can be processed at a time. An example of this was to remember telephone numbers in chunks of numbers, instead of individual numbers. Miller's findings are rather dated, and have been criticized in many papers since. In particular, critics have pointed out that differences in the type of information can affect the amount that can be stored. In many cases, the number has been found to be even lower than Miller stated. However, the general idea of processing information in chunks is quite similar to the SOM model in this thesis; the companies are viewed as members of particular clusters, each of which exhibit a particular combination of financial ratios. From this viewpoint, the SOM could be viewed as a complexity reducing tool, potentially capable of reducing the complexity reported by managers in Table 3.4.

3.3 Conclusions

The purpose of this chapter was to assess the current state of the art in financial benchmarking in Finnish, publicly-noted companies. Also, managers' satisfaction with current methods was assessed, in order to determine if there is a need for alternative methods.

Firstly, the key concepts and competing methods were briefly identified and presented. Then, the results of a survey of business intelligence managers in

Finnish publicly-noted companies, conducted in order to assess the current state of the art in financial benchmarking, were presented.

In conclusion, one can state that although the surveyed managers are quite familiar with IT, and use common IT tools daily, very few advanced methods are used to process financial competitor information. In addition, the most commonly used source of financial information is newspapers. Managers are primarily using single ratio approaches instead of multiple ratio approaches. The SOM is thus potentially a valuable addition to currently used methods.

The format of information and ease of use were obviously the most lacking factors in current methods in use. Generally speaking, managers were fairly satisfied with all other factors in current methods. However, it is clear that there is room for improvement in their current methods, and the hypothesis is that the self-organizing map could provide additional value to the managers. The primary strength of the SOM, its visualization properties, is potentially a particularly valuable addition, as this area is most lacking in current methods.

Although managers are divided concerning information overload, it is obvious that there is a large group of managers in need of tools to help them deal with information overload. In addition, the complexity of the competitive environment was viewed as high.

The high complexity of the competitive environment, the lack of advanced tools in use, and managers' varying degree of satisfaction with their current methods, indicates that there is room for improvement in the methods used for financial benchmarking in Finnish publicly noted companies. In this thesis, the suitability of the SOM for providing a solution to these problems will be assessed.

The results also showed that 36.84% of managers had previously heard of the self-organizing map, and 39.47% expressed a potential interest in participating in the second phase of the evaluation. Thus, roughly a third of the managers had heard of the SOM but none of the examples mentioned were from finance, and were typically in the form of ongoing academic research.

4 BENCHMARKING

This chapter will first briefly discuss the history of benchmarking, as well as the different forms of benchmarking available. Then, the benchmarking process will be defined. Finally, the form of benchmarking used in this dissertation will be presented.

4.1 Definition

Merriam-Webster defines benchmarking as: "The study of a competitor's product or business practices in order to improve the performance of one's own company"[7]. The Xerox Corporation, one of the pioneers in benchmarking, uses the following definition: "Benchmarking is the continuous process of measuring our products, services and practices against the toughest competitors recognized as industry leaders" (Gustafsson 1992, p.10).

Benchmarking is the process of comparing the activities of one company to those of another, using *quantitative* or *qualitative*[8] measures, in order to discover ways in which effectiveness could be increased (Karlöf and Östblom 1994). The focus of benchmarking is not only on *what* the competitors do, but *how* they do it. Benchmarking is about discovering new processes that can be used to increase the own company's effectiveness, i.e. learning from the competitors (Gustafsson 1992, p.9). Karlöf (1997, p.14) calls it inspiration from, not imitation of, a competitor. Depending upon the form of partnership in the benchmarking process, it is also fully possible for all companies involved to benefit from benchmarking.

4.2 Background

Benchmarking is not an entirely new concept. Various forms of comparing standards of performance have in fact existed throughout history. An example of an early form of benchmarking is the master craftsmen's associations of the Middle Ages. These not only served to protect the member's livelihoods, but also to guarantee a standard of quality of the products produced. For example, in

---

[7] According to Merriam-Webster's Collegiate Dictionary, at http://www.m-w.com/dictionary.htm

[8] Quantitative measures are numerical data, while qualitative measures are descriptive data. Quantitative measures can be further divided into financial and non-financial measures (McNair and Liebfried 1992, p.167)

England wardens were appointed to ensure that the work was of an acceptable quality. (Bendell et al. 1998, pp.45-46)

The first to actually implement benchmarking in a form comparable to today were the Japanese. After World War II, the Japanese economy was in ruins, and most industry had to be rebuilt from scratch. In the 1950s, Japanese industry leaders paid visits to western organizations, carefully noting the manufacturing processes used. These organizational processes were then compared to each other, and the best processes were selected. The Japanese call this process *dantotsu*, which means to strive to be "the best of the best" (Camp 1993, p.13). The Japanese were very successful in their adoption of the best business practices available and Japanese companies quickly grew to compete with, and beat, western companies. (Bendell et al. 1998, p.66)

The first western company to implement benchmarking was the Xerox Corporation, in the late seventies. Xerox was facing heavy competition from companies who were able to sell their products much more cheaply, in fact at prices much lower than those at which Xerox could manufacture them. In 1979, Xerox implemented a program, coined *competitive benchmarking*, in order to understand how this was possible, and what the competitors were doing differently. Initially, only a few divisions were involved, but by 1981 the entire corporation was involved in benchmarking (Camp 1993, p.17). The Xerox strategy was an immediate success (Gustafsson 1992, p.57; Bendell et al. 1998, pp.66-71). Today, benchmarking is a very popular concept in both research and practice. For example, Dattakumar and Jagadeesh (2003) list more than 350 publications on the topic by June 2002.

4.3 Different Methods of Benchmarking

There are several methods of benchmarking. The type of benchmarking method applied depends upon the goals of the benchmarking process. Bendell et al. (1998, pp.82-84) divide benchmarking methods into four groups. These are: *internal*, *competitor*, *functional*, and *generic* benchmarking.

Internal benchmarking is a method in which the performance of one part of an organization is compared to other parts. This form of benchmarking is very common and easy to arrange, but is unlikely to produce results that are world best practice. (Bendell et al. 1998, pp.82-84)

Competitor benchmarking is much more difficult to implement than internal benchmarking, but also much more likely to achieve best practice within the industry. In this case competitors are used as benchmarking partners. The problem is that many companies do not for reasons of confidentiality want to reveal all of the necessary information. One must remember that even though companies might be benchmarking partners, they are still often competitors. This form of benchmarking can, of course, also be implemented without direct contact with the benchmarking partner. (Bendell et al. 1998, pp.82-84)

Functional benchmarking involves making comparisons against non-competitor organizations, which are good at a particular activity that a company is interested in. Examples could be warehousing, procurement, etc. One advantage with this form of benchmarking is that confidentiality is not usually an issue, as the involved companies are not competitors. This could also lead to novel practice in the industry, as knowledge from a different industry is imported. (Bendell et al. 1998, pp.82-84)

Generic benchmarking is the final method of benchmarking. Generic benchmarking is the most extensive form of benchmarking, and involves benchmarking across several, not necessarily related, industries. The results of this can be very innovative, and are most likely to create breakthroughs. However, this is also the most challenging to implement, and the results can be very difficult to put into practice. (Bendell et al. 1998, pp.82-84)

In addition, benchmarking can be divided into three types depending upon the goal of the benchmarking: *performance benchmarking, process benchmarking,* and *strategic benchmarking* (Ahmed and Rafiq 1998; Bhutta and Huq 1999). Performance benchmarking is the comparison of different companies according to defined performance measures. Process benchmarking involves the comparison of methods and processes in different companies in order to improve own processes. Strategic benchmarking can be used when an effort to change the strategic direction of the company is being made. The suitability of each benchmarking method for a particular goal is illustrated in Table 4.1.

	Internal benchmarking	Competitor benchmarking	Functional benchmarking	Generic benchmarking
Performance benchmarking	Medium	High	Medium	Low
Process benchmarking	Medium	Low	High	High
Strategic benchmarking	Low	High	Low	Low

Table 4.1. Suitability of benchmarking methods for different goals (Bhutta and Huq 1999), originally adapted from (McNair and Liebfried 1992).

4.4 The Benchmarking Process

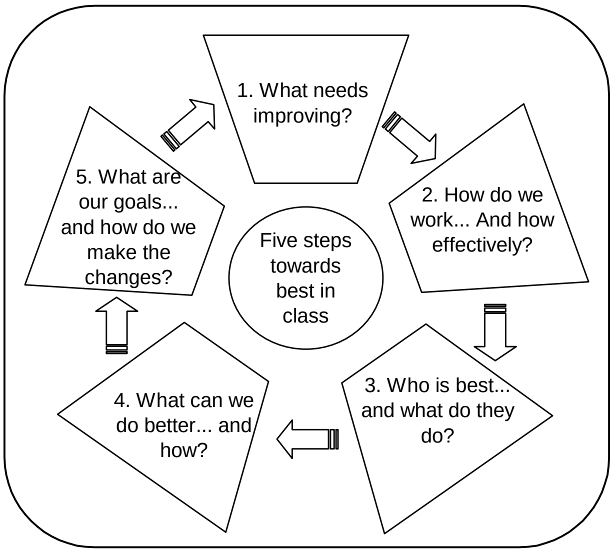


Figure 4.1. The benchmarking process. (Gustafsson 1992, p.28)

One model of the benchmarking process is illustrated in Figure 4.1. There are many definitions of how many steps are needed in the benchmarking process. For example, Bendell et al. (1998, pp.78-86) propose a four-step model, Camp (1989) and Bhutta and Huq (1999) propose a five-step model, and Xerox uses a ten-step process (Bendell et al. 1998, p.67). In essence, most models are very similar, varying only the amount of steps used. Gustafsson (1992, pp.28-44) proposes a five-step process:

1. Decide what needs to be improved upon.
2. Map the own processes, and measures of effectiveness.
3. Find examples, establish contacts and compare performance.
4. Analyze differences.
5. Establish challenging goals and complete the improvements.

In the first step, the company must decide what to benchmark. This could be any aspect of a company's activities, or a mix of several. Examples of these are quality, productivity, customer relations, financial performance, etc. Then, the method of benchmarking (internal, competitor, functional, or generic benchmarking) is chosen. (Gustafsson 1992, pp.29-31)

In the second step, own processes are mapped and described, and the measures of performance to be used are chosen. The purpose of this step is to describe and understand the company's own way of performing a particular activity, and to choose the way in which performance is to be measured. (Gustafsson 1992, pp.32-36)

In the third step, the benchmarking partners are chosen. The choice of partners will depend upon the method of benchmarking (internal, competitor, functional, or generic benchmarking). Direct contact is not necessary when comparisons are performed using external measures, for example, using financial data from annual reports. (Gustafsson 1992, pp.36-40)

The fourth step involves comparing and analyzing the differences in processes. The reasons for the differences are analyzed, and suggestions for improvements are made. (Gustafsson 1992, pp.40-42)

Finally, in the fifth step, new goals are set, and the implementation of changes is initiated. As benchmarking is a continuous process, the process is restarted and once again, processes that need improving upon are selected. (Gustafsson 1992, pp.42-44)

4.5 Financial Competitor Benchmarking

Financial benchmarking is commonly used for comparing the financial performance of competing companies. Financial benchmarking is an example of external (competitor) benchmarking for performance comparison (see Table 4.1), where the performance measures consist of quantitative financial ratios (Bendell et al. 1998; Camp 1989, pp.62-63; Reider 2000, pp.146-147). Companies' financial ratios are commonly calculated on the basis of figures given in financial statements. This type of competitor analysis has been proven to be important, and is performed regularly in many companies (Guilding 1999).

Benchmarking in general deals with improving the underlying processes in a company by comparing internal processes to those of successful benchmarking partners (Zairi 1994). Financial benchmarking does not specifically address the improvement of processes. It can, however, be seen as the first step in determining the competitors to benchmark against, in order to identify industry best practice (Camp 1989, pp.66-68). When the best performing company has been identified, benchmarking is used to discover the real reason for the level of performance, and to analyze how this level of performance might be achieved (Camp 1989, pp.17-19, 66-68; Reider 2000, p.31, pp.145-146, 177-179).

4.6 The Benchmarking Process in this Study

This particular study is an example of competitor financial benchmarking using quantitative data. The goal of the benchmarking process is performance comparison (Table 4.1). Different companies that are competitors within the same industry are benchmarked against each other using various financial ratios. The information used in the study is taken from companies' annual reports. This benchmarking process differs in some ways from the traditional definition of benchmarking. Firstly, the study is from an outsider's view, and not from the view of a company, in order to produce an objective study. Secondly, only publicly available information is used, since no personal contact is made with the individual companies. Finally, benchmarking is not in this case used to achieve better performance, but simply to objectively compare the performance of different companies from different geographical areas. It can also be used to identify where the "state of the art"-level currently is. Also, the study can be seen as an alternative to traditional consulting company reports, in which information is presented as a traditional ranking.

In this study, the benchmarking process is as follows (Figure 4.1). In Chapter one, the purpose of the benchmarking is identified, i.e. the question "What needs improving?" is answered (Step 1). The performance measures are selected in Chapter five and the benchmarking partners (objects) are identified in Chapter seven (Step 2). The gathering of data (Step 3) is also discussed in Chapter seven. The actual benchmarking and analysis of the companies is presented in Chapter eight (Step 4). Step five, implementation of the results, is beyond the scope of this dissertation.

5 FINANCIAL RATIO ANALYSIS (FRA)

The most common form of financial benchmarking involves the measuring of a company's financial performance according to a selection of financial ratios. This chapter introduces financial ratio analysis (FRA), which deals with the selection and calculation of financial ratios. In this chapter, the choice of financial ratios will be discussed and motivated. The chapter begins with a brief discussion of the classification of financial ratios. Problems due to differing international accounting practices, and how they affect financial ratios, are then discussed. Finally, the ratios used in this dissertation, and the motivations for these choices, are presented.

5.1 Introduction

Traditionally, financial ratios have been used in order to make items comparable across firms and over time by adjusting for size (Salmi and Martikainen 1994), and to measure some specific aspect of financial performance. Financial ratios are divided into classes depending upon the items that they measure. There are many different classifications, but the traditional *a priori* classification includes the following four classes: *profitability ratios, short-term solvency (liquidity) ratios, long-term solvency ratios,* and *efficiency (turnover) ratios* (Foster 1978, p.28; Lev 1974, p.12).

Profitability ratios measure the ability of a company to generate revenues in excess of expenses relative to the capital used (Foster 1978, p.33). Examples of profitability ratios include *net margin, operating margin, return on equity,* and *return on total assets.*

Liquidity ratios measure the short-term payment ability of a company. Inventory, raw materials, accounts receivable, cash on hand, etc., are considered liquid assets, which can quickly be converted into cash. Liquidity ratios measure how much of the company's short-term liabilities (current liabilities) can be covered using these liquid assets. Commonly used measures include *current ratio* and *quick ratio* (Bertoneche and Knight 2001, pp.86-87).

Long-term solvency ratios measure a company's ability to meet its long-term financial obligations. Solvency ratios can include both traditional indebtedness ratios, such as *equity to capital,* and measures that include profit, such as *interest coverage*. In some classifications (e.g. Lainez and Callao 2000), indebtedness and solvency are separate classes.

Efficiency ratios, or turnover ratios, measure a company's ability to efficiently use its assets. Each efficiency ratio deals with some aspect of asset efficiency, such as *asset turnover, inventory turnover,* or *receivables turnover*.

In addition, financial ratios are also classified based on the values that they employ. *Dynamic ratios* employ data from the income statement, while *static ratios* employ data from the balance sheet (Salmi and Martikainen 1994). In addition, there are so-called *mixed ratios*, which employ data from both the income and balance sheets.

5.2 Accounting differences

Any international financial analyst must remember that accounting differences exist between different countries, and that these differences can greatly influence the reliability of any analysis (Lainez and Callao 2000). This is, of course, also true in financial benchmarking, as the results are meaningless unless the analyst is aware of the rules and regulations that accountants follow (McNair and Liebfried 1992, p.168).

There are a number of different reasons for the existence of international accounting differences. These are primarily *the external environment and culture, legal systems, prevailing financing sources, taxation, the development of the accounting profession, and inflation* (Choi et al. 2002, pp.42-45; Nobes and Parker 2002, pp.20-30). The most important of these factors are the prevailing financing sources and the legal system (Nobes and Parker 2002, p.18), as these define who the reporting targets, as well as the form of the reporting. On the one hand, in an equity-based market, the ownership is public and the main principle is to help the shareholders assess the performance of management (i.e. fairness) (Nobes and Parker 2002, p.23). On the other hand, in a credit-based market, ownership of companies is primarily by large credit institutions, such as banks. In these cases, the emphasis is on creditor protection through the use of conservative accounting (Choi et al. 2002, p.43). Along largely the same lines, accounting practices are determined by the prevailing legal system. In the *common law system* (or case law), laws are written to answer a specific case, with no attempt to write all-encompassing laws. Instead, legal cases (precedents) are used to expand and interpret existing laws. Thus, judgment becomes an important part of the common law system. In the *codified law system* (or Roman law) justice and morality are central, and codified laws tend to be all-embracing sets of requirements and procedures. In codified law countries, accounting rules and procedures are generally included in the law, and tend to be highly prescriptive and procedural. On the other hand, in common law systems, accounting rules are generally created by organizations representing the accounting profession, generally making them more adaptive and innovative (Choi et al. 2002, p.43;

Nobes and Parker 2002, pp.20-21). Finally, taxation has a great influence on accounting differences. In some countries, typically those using codified law, taxation influences accounting laws as taxation is based directly upon the financial statements. In other countries, separate statements are made for taxation purposes. In countries with taxation based upon the financial statements, there are, of course, incentives to reduce profit with the help of income smoothing measures, in order to reduce the amount of tax paid. This can lead to great differences in financial figures (Choi et al. 2002, p.43; Nobes and Parker 2002, pp.25-27).

Generally speaking, the accounting world can be divided into two spheres of influence: the *Anglo-Saxon* and the *Continental European* (Nobes and Parker 2000). Nobes and Parker (2002) also refer to the Anglo-Saxon model as the *Strong Equity accounting system*, and Continental European model as the *Weak Equity accounting system*, based upon the principal corporate ownership structure in the different countries (Nobes and Parker 2000, p.62). This influences the reporting in the way discussed above, i.e. fairness vs. conservative accounting. In addition, since accounting rules are based on common law instead of codified Roman law, the Anglo-Saxon model contains less specific rules for how financial statements should be presented (Nobes and Parker 2000, p.19; Lehtinen 1996, p.22). Finally, in the Anglo-Saxon model of accounting, while having a small effect, corporate taxation does not determine the contents of the financial accounts. The Anglo-Saxon model is prevalent in English-speaking countries, i.e. those with former ties to the British Empire. This means that the prevailing accounting laws in these countries are very similar. Examples of such countries are the UK, USA, Canada, Australia, Ireland, and South Africa (Nobes and Parker 1991, pp.123-125). There has been some discussion as to the actual existence of a homogeneous Anglo-Saxon accounting model (Alexander and Archer 2001; response by Nobes 2003), but it is generally accepted in the literature.

While the Anglo-Saxon model is considered fairly homogenous, the Continental European model is much less so, and incorporates a large number of very different systems. Continental European accounting involves a heavier influence on taxation laws, since these countries regularly use income statements as the basis for corporate taxation (Nobes and Parker 2000). This, of course, implies that different methods of income smoothing and adjustments exist in these countries. Anglo-Saxon countries generally do not have these, since separate statements are used to calculate corporate taxes in these countries. In addition, with their heavy influence on ideas of justice and morality, Continental European methods rely on much more detailed company law or commercial codes (Lehtinen 1996, p.22). This is typical for countries employing legal systems based on Codified Roman Law (Nobes and Parker 2000, p.19).

The *International Accounting Standards* (IAS) is an international effort towards harmonization of international accounting practices. The ultimate goal of IAS is to achieve one comprehensive, world-wide accounting framework, which should make comparison of international accounts possible. In many countries, efforts have been made to adapt national accounting practices towards the new IAS regulations, but in many cases accounts are still far from comparable. However, as a result of this, accounting differences have decreased between many countries. IAS is discussed in more detail in Section 5.3.

While some companies do report the figures according to IAS in addition to national GAAP, all companies do not. In addition, these figures are not always as detailed, nor even comparable, as IAS has been continuously evolving (see Section 5.3). Therefore, international accounting differences must be assessed, and ratios that minimize these differences must be selected. In the literature, there are a number of important areas in which accounting practices differ significantly from country to country: *depreciation method, valuation of inventory, accounting for leases, research and development costs, accounting for goodwill, valuation of assets, provision and reserves,* and *deferred taxation.*

Depreciation Method

Differing depreciation methods are one of the most difficult problems in international accounting. Depreciation can be used to lower a company's result, thereby decreasing the amount of tax it has to pay. This system is in use, for example, in Finland, France, Germany, Japan, and Sweden, and most other countries with accounting systems based upon the Continental European model. In many other countries, for example, the USA, UK, and the Netherlands, taxation is calculated separately, and depreciation takes place at a fixed annual rate. Depreciation methods have the strongest effect on profitability and equity ratios, while having little effect on liquidity and efficiency ratios. (Lehtinen 1996, pp.53-54; Choi et al. 2002, pp.65-97; Nobes and Parker 2002)

Valuation of Inventory

Valuation of inventory is another problem. In different countries, different costs are included in the inventory value. For example, in Finland fixed overhead costs can only under certain conditions be included in the inventory value. In some countries, these conditions are not as strict, while in other countries fixed overhead costs cannot be included at all. Another concern is that for instance in the USA, the LIFO[9] (Last in, first out) principle is commonly used when valuating the inventory, whereas the prevailing principle in most other countries

[9] LIFO means that the next unit consumed / sold is valued at the cost of the last unit purchased / produced. FIFO implies that the next unit consumed / sold is valued at the cost of the first unit purchased / produced

is FIFO (first in, first out). This can lead to considerable differences in the value of the inventory. When prices are rising, a LIFO inventory will be valued higher than a FIFO inventory, and when prices are sinking, the opposite will occur. This will, of course, affect the profit as the cost of inventory changes. Valuation of inventory will have a high effect on liquidity and efficiency ratios, and also a small effect on profitability ratios. (Lehtinen 1996, pp.55-56; Choi et al. 2002, pp.65-97; Nobes and Parker 2002, p.75)

Accounting for Leases

There are two kinds of leases: *finance leases* and *operating leases*. Finance leases are a form of financing in which the ownership of an asset often remains with the lender, but the economic risk is borne by the borrower. With operating leases on the other hand, the economic risk is usually borne by the lender. Operating leases never show up in the balance sheet, and are in all countries treated as renting operations. Finance leases are treated differently in different countries. For example, in Finland and Germany finance leases may not be capitalized. In countries like the UK, USA, and Sweden, finance leases are capitalized and depreciated like other assets. With the increasing use of leasing as a form of financing, the effects of these differences can be expected to increase in the future. Leasing practices may have an effect on many different financial ratios, mostly on profitability and solvency ratios. (Lehtinen 1996, pp.56-57; Choi et al. 2002, pp.65-97)

Research and Development Costs

In some countries, like the USA, research and development (R&D) costs must be treated as immediate expenses. In many countries, under certain conditions, R&D costs may instead be capitalized, such as France, the Netherlands, and the UK. On the other hand, according to International Accounting Standards (IAS), R&D costs may never be capitalized. R&D cost capitalization will lead to effects on several financial ratios, especially on profitability ratios. (Lehtinen 1996, pp.57-58; Choi et al. 2002, pp.65-97)

Accounting for Goodwill

Goodwill can be accounted for in several ways:

1. It can be capitalized under intangible assets and amortized,
2. It can be charged directly to the income statement, or
3. It can be charged directly to the reserves.

If goodwill is capitalized and amortized (i.e. depreciated as an asset), it will affect both the income statement and the balance sheet. This is common practice in, for example, Finland and Germany. These countries also have different depreciation periods for goodwill. In some countries, like Japan, it is only possible to charge goodwill directly to the income statement. Accounting for goodwill can affect most financial ratios, with the exception of efficiency ratios. Profitability ratios are especially sensitive to accounting for goodwill. (Lehtinen 1996, pp.58-59; Choi et al. 2002, pp.65-97)

Valuation of Assets

The valuation basis is a source of differing practices. In most countries, historical costs are used as the basis for valuation of fixed costs. Revaluation of assets is not allowed at all in some countries, notably in the US and Germany. In some countries revaluation is allowed, but does not affect the profit at all, only the balance sheet. Finland is an example of a country where this principle is used. Other countries, like the Netherlands and the UK, allow revaluations that affect both the balance sheet and income statement. These differences might affect both profitability and solvency ratios. (Lehtinen 1996, p.59; Choi et al. 2002, pp.65-97; Nobes and Parker 2002, pp.40-41)

Provisions and Reserves[10]

Provisions are a form of income smoothing, used in countries where the corporate tax is based on the income statement, for example, in Germany. Reserves, on the other hand, are a way of dealing with uncertainties concerning future revenues and losses. Reserves are more commonly used in countries where the corporate tax is not based on the income statement, such as in the USA and the UK. Provisions and reserves affect profitability and equity ratios. (Lehtinen 1996, pp.59-60; Choi et al. 2002, pp.65-97; Nobes 2002, pp.38-40, 114)

Deferred Taxation

Finally, differences in taxation policies in years of negative income can cause ratios to become incomparable. In many countries, a lower tax rate during years of negative profit can be balanced with correspondingly higher taxes during the following years. However, the practices regulating this process differ greatly, and in some countries, like Germany and Japan, deferred taxes are only allowed under certain circumstances. Deferred taxes can have a strong effect on many ratios, since they affect net profit and liabilities. (Lehtinen 1996, p.60; Choi et al. 2002, pp.65-97; Nobes and Parker 2002)

[10] There is some controversy concerning the definition of the words provision and reserve. In US English, they are generally used interchangeably. In this case, the UK English definition, as implied above, will be used. (Nobes and Parker, 2002, p.38)

5.3 International Accounting Standards (IAS)

The diversity in international accounting practices has led to a number of international harmonization attempts. The most important of these are the *International Accounting Standards* (IAS). IAS is a set of standards stating how different transactions and events should be reflected in financial statements. The standard is maintained by the *International Accounting Standards Board* (IASB), the board of the *International Accounting Standards Committee* (IASC). The IASC has no formal jurisdiction to enforce these standards, but many countries require their publicly-noted companies to present their financial statements in accordance with IAS. The IASC was founded in 1973, and is the most important and most successful international body for accounting harmonization, with accounting body members from over 100 countries in 2000 (Nobes and Parker 2000, p.69).

Generally, the IAS more or less closely follows and compromises between US and UK GAAP (generally accepted accounting procedures) (Nobes and Parker 2000). However, the US notably still requires foreign companies noted on the New York Stock Exchange (NYSE) to publish their accounts in US GAAP (Harris and Muller III 1999). IAS is not sufficient. Street et al. (2000), however, found that differences in US GAAP and IAS accounts published by foreign companies noted on the NYSE were not significant. In the European Union, all publicly-noted companies will be required to post IAS-compatible accounts no later than 2005 (Street 2002).

El-Gazzar et al. (1999) found that companies operating on international financial markets voluntarily adopt IAS for a number of reasons, including increased access to foreign capital and foreign markets, improved customer recognition, and reduced political costs of doing business abroad. These companies are more likely to voluntarily disclose higher levels of investor-oriented information. For example, many internationally active Finnish companies, including UPM-Kymmene, have already published both Finnish GAAP and IAS statements for several years, and will be fully adopting IAS from January 1, 2004.

However, Street (2002) benchmarked national GAAPs in 62 countries against IAS, and found significant differences. The only countries entirely compliant were Cyprus, Kenya, and Romania, while South Africa, Peru, and Mexico had the least differences. Most problematic were Russia, Switzerland, Spain, Greece and Luxemburg. Another problem is actual - as opposed to stated – compliance with IAS in individual companies (Street and Bryant 2000; Street et al. 2000; Taylor and Jones 1999; Street et al. 1999). It is clear that IAS continues to be a challenge in international accounting, and accounting harmonization remains elusive. However, as El-Gazzar et al. (1999) note, we are not too far from a set of worldwide accounting principles. IAS will certainly, in the long run, increase international financial comparability.

5.4 Choosing suitable ratios for international financial benchmarking

It is obvious that IAS is not an option for dealing with international accounting differences yet, as data as old as 1994 are included. One way to reduce the problems brought about by international accounting differences would be to create a model that takes into account at least some of these differences. A number of authors have proposed different models that deal with the standardization of financial reports. For example, one way would be to calculate an adjustment index (Whittington 2000) to adjust for differences in accounting practices. Another, more time consuming way is to create a model that actually converts statements between two or more accounting systems (e.g. Speidell and Bavishi 1992; cited in Choi et al. 2002, p. 358). However, creating and using these kinds of models is very difficult and time-consuming. It was, therefore, preferable to try to find financial ratios that were least affected by accounting differences. This can be done by judging the *reliability* and *validity* of the individual ratios, as was done by Lehtinen (1996).

5.4.1. Reliability and Validity

The accuracy of a financial ratio can be rated according to two criteria: *validity* and *reliability*. A financial ratio's validity refers to its capability to measure what it is supposed to measure. For example, profitability is often defined as the ability of a firm to generate revenues in excess of expenses relative to the capital used. This definition implies that a correct measure of profitability requires that both profit and capital be taken into account. Thus, Operating Margin has a low theoretical validity while Return on Assets has a high validity.

The reliability of a financial ratio implies how sensitive it is to differences in accounting principles. A ratio that can easily be manipulated by a company's choice of accounting policy cannot be considered reliable.

In this thesis, the choice of ratios is based on an empirical study by Lehtinen (1996), in which both the theoretical and empirical validity and reliability of a number of financial ratios were assessed. Lehtinen's study consisted of two parts. In the first part, the author studied the theoretical validity and reliability of a number of financial ratios. In the second part, an empirical study was carried out to find a number of ratios best suited for international comparisons.

The dataset consisted of 16 financial ratios, calculated for 1,230 companies from Finland, Sweden, the U.K., the Netherlands, Italy, France, Japan, and Germany.

The ratios were calculated based on information from income statements and balance sheets. First, Lehtinen (1996) used factor analysis and transformation analysis to determine if the different ratios are measures of the same economic property in the different countries. The findings indicated that profitability and efficiency ratios were connected in all of the studied countries. Solvency and liquidity ratios were not as easy to connect. Thus, the empirical validity of profitability and efficiency ratios is high (Lehtinen 1996, pp.102-103). The empirical validity of solvency and liquidity ratios, on the other hand, was low.

The author then used analysis of variance to determine the level differences between the ratios of the different countries. Lehtinen concluded that profitability ratios are influenced by international accounting differences, causing significant differences in country ratios. Liquidity, solvency, and efficiency ratios, however, differ little across countries (Lehtinen 1996, p.125). Lehtinen further concludes that “the validity of financial ratios is a more important factor than the reliability” (Lehtinen 1996, p.131). Thus, analysts are encouraged to choose ratios based on their validity. Finally, Lehtinen proposes the following ratios for international comparisons: *Operating Margin, Return on Total Assets, Quick Ratio, Defensive Interval, Equity to Capital, Interest Coverage*, and a combination of all of the presented efficiency ratios.

Lehtinen’s (1996) findings are somewhat different from those of Lainez and Callao (2000). For their paper, Lainez and Callao compared a number of posts from annual reports completed according to different international accounting practices. The differences were then compared for statistical significance. The authors indicate that the ratios most affected by international accounting differences are solvency and liquidity ratios. Profitability ratios on the other hand did not seem to be heavily influenced. The findings indicate that the most significant differences are due to three items: *valuation of fixed assets, recognition of goodwill*, and *valuation of exchange losses*.

Lehtinen’s (1996) study was used as the basis when choosing the ratios to be used in this dissertation, since it includes many of the countries relevant to this study, including Canada and the Scandinavian[11] countries. Although the study is fairly dated, it can be assumed that accounting differences have been reduced in the years since the study, instead of the opposite. A priori, one can therefore assume that there exist larger differences among the ratios calculated based on old data from 1995 than among those based on more recent figures, making the use of a rigid framework, such as Lehtinen’s, feasible.

A recurring problem during the experiment has been the different definitions of financial ratios used by companies in their annual reports. Whereas ratios like Operating Margin and Current Ratio have been identical, some ratios, like Return

[11] In this study, Finland is included in the Scandinavian countries

on Equity and Return on Total Assets, have not. This means that it has not been possible to use the values provided by the companies, and these have had to be manually calculated. In addition, most of the chosen ratios were not provided in the annual reports at all. Therefore, Foster's (1978) and Lev's (1974) definitions have been used, and all ratios have been manually calculated.

5.5 Chosen Ratios

The ratios to be used in this dissertation will now be presented, along with the motivations for their selection. The number of ratios chosen was limited to seven, as this provides room for a very broad selection of ratios.

5.5.1. Profitability Ratios

The proportion of profitability ratios in this experiment may at first glance seem large, but profitability can be considered to be the most important measure of a company's performance. This is because in the long run, a company must remain profitable in order to survive, and also in order to maintain both good liquidity and solvency. In addition, the selection of profitability ratios is very broad, with a large emphasis on return on invested capital. By including many profitability ratios, the model can be made to prioritize companies showing high profitability.

Equation 5.1

$$Operating\ Margin = \frac{Operating\ Profit}{Sales} \times 100\,.$$

Operating Margin was chosen as a ratio for two reasons: it is simple to use and it is relatively unlikely to be heavily influenced by accounting practices. It is also a very commonly used profitability measure.

Equation 5.2

$$Return\ on\ Equity = \frac{Net\ Income}{((Share\ Capital + Retained\ Earnings)average)} \times 100\,.$$

Equation 5.3 (Foster 1978, p.33)

$$Return\ on\ Total\ Assets = \frac{Total\ Income + Interest\ Expense}{Total\ Assets\,(average)} \times 100\,.$$

The problem with Operating Margin is its low theoretical validity, as it does not take into account the capital of a company, which is a requirement for accurate measuring of profitability (Lehtinen 1996, pp.50-51). This is why Return on Equity (ROE) and Return on Total Assets (ROTA) have been included. There are many different definitions of both Return on Equity and Return on Total Assets, as can be seen by those provided by different companies in their annual reports. The formulas provided by Foster (1978, p.33; also used by Lehtinen 1996, pp.62-63) have been used as standard. Return on Total Assets was the other profitability ratio proposed by Lehtinen. Return on Equity was included because it is perhaps the most commonly used profitability measure (Bertoneche and Knight 2001, p.79).

5.5.2. Liquidity Ratios

Equation 5.4 (Foster 1978, p.33)

$$Quick\ Ratio = \frac{Current\ Assets - Inventory}{Current\ Liabilities}\ .$$

Quick Ratio is a measure of the company's liquidity, or short-term payment ability. Quick Ratio, as opposed to the more commonly used Current Ratio, discounts the current inventory from current assets. Most of the companies in this study had rather large inventories, and these inventories are in some ways not nearly as liquid or stable as other current assets. Therefore, the Quick Ratio is in this case a better reflection of the company's short-term payment ability. The advantage of using Quick Ratio is its reliability in an international context. The problem is its low theoretical validity. Lehtinen suggested the use of Defensive Interval as well. However, while this ratio would be both reliable and valid, it would be much more complicated to calculate, so it has been left out. Also, this avoided placing too much emphasis on liquidity in this study.

5.5.3. Solvency Ratios

Equation 5.5

$$Equity\ to\ Capital = \frac{Share\ Capital + Retained\ Earnings}{Total\ Assets} \times 100\ .$$ [12]

[12] A mistake in the data preprocessing stage caused the equity to capital ratio to be incorrectly calculated. The denominator was calculated as an average over the entire year, instead of a snapshot at the end of the year. The result is overall higher values in equity to capital. However, for companies whose balance totals change significantly during the year, for example due to mergers, the effect will be larger. The mistake was noticed after

There are some disadvantages with using Equity to Capital (primarily a low reliability), but it is a very commonly used ratio that is fairly easy to calculate, and its use is also suggested by Lehtinen.

Equation 5.6

$$Interest\,Coverage = \frac{Interest\ Expense + Income\,Tax + Net\ Income}{Interest\ Expense}.$$

Equity to capital is a static ratio, meaning that it does not take into account the cash flow of the company (Foster 1978, p.31). Interest coverage has been included to provide a ratio that offsets this problem. Interest Coverage also benefits from a relatively high validity. According to the empirical study, Interest Coverage is a good solvency ratio.

5.5.4. Efficiency Ratios

Equation 5.7

$$Receivables\,Turnover = \frac{Net\ Sales}{Accounts\ Receivable\,(average)}.$$

Receivables Turnover is possibly the best efficiency ratio where both reliability and validity are concerned. It was, therefore, chosen as the sole efficiency ratio. Total Assets Turnover and Working Capital Turnover might have been interesting ratios to use, but their results are much more heavily influenced by accounting differences. While Lehtinen (1996) suggests that all efficiency ratios should be used at once, this would have placed too much emphasis on efficiency. Therefore, Receivables Turnover was chosen for its good validity and reliability, as well as its ease of use.

The final choice of ratios is summarized in Table 5.1.

---

the entire evaluation was performed, indicating that the effects are not unreasonably large.

Class	Ratios
Profitability ratios	1. Operating margin 2 Return on total assets 3 Return on equity
Liquidity ratios	4 Quick Ratio
Solvency ratios	5 Equity to capital 6 Interest coverage
Efficiency ratios	7 Receivables turnover

Table 5.1. Summary of financial ratios used in this study.

6 NEURAL NETWORKS AND SELF-ORGANIZING MAPS

Neural networks are based upon algorithms that seek to emulate the human brain’s method of solving problems. The following sections deal with the concept of neural networks, with an emphasis on self-organizing maps.

6.1 Background

Alexander Bain (1818-1903) proposed the first neural network in 1873. His proposal was based on recent advances in neuroanatomy. Well before any models of the neuron were firmly established, Bain presented proposed schemes for how neural connections might work. Actually, Bain proposed an early version of the principles proposed by Hebb much later. However, Bain’s research was well ahead of its time, and his research did not become commonly acknowledged until much later. (Olmsted 1998; Wade 2001)

McCulloch and Pitts (1943) introduced the notion of threshold logic, and added several input lines instead of the two used until then. They therefore proposed that the output of the neuron be analog, instead of the binary output in use at the time. They also introduced the concept of weighted connections, thereby presenting the first adaptive networks. (Olmsted 1998; McCulloch and Pitts 1943)

In 1949, a psychologist by the name of D. O. Hebb proposed a learning law (Hebb 1949) that made possible the learning algorithms of today. During the 1950’s and 60’s, this model was used by researchers to create the first artificial neural networks. These networks were called *perceptrons* (Olmsted 1998). This led to an explosive growth in the amount of research on the topic. However, it soon turned out that these simple one-layer neural networks were incapable of solving even simple problems. (Wasserman 1989, p.4)

The setbacks suffered using perceptron networks inspired researchers to modify their models. This led to the first multilayered network, called the *cognitron* network, by Fukushima in 1975. Finally, in 1986, the *back-propagation network* was developed by Rumelhart and McClelland (1986).

In 1972, Prof. Teuvo Kohonen took a slightly different approach to neural networks. He based his work on the notion that memory may be holographic in nature. Kohonen suggested that the brain consists of a number of *receptive fields,* which each respond to different stimuli (Kohonen 1972). Thus, the brain could be

seen as a number of *ordered feature maps*, grouped according to similarities. (Kohonen 1989, pp.6-7)
This theory led Kohonen to introduce the self-organizing map in 1981. (Kohonen 1990).

6.2 Definition

A neural network is built up in a pattern similar to that of the human nervous system. The human nervous system consists of billions of *neurons*, interconnected by a network of *synapses*. The neurons can receive information from the outside world at several different points in the network. This information is called *stimuli*. The stimuli travel through the network, generating different responses in different neurons, which in turn send, or *fire*, new internal signals to neighboring neurons. These signals can be of various strengths, depending upon their importance. The signal will cause a reaction in an individual neuron, either *exciting* or *inhibiting* it. If it is excited, a neuron will pass the signal to neighboring neurons, but if it is inhibited, it will not. This eventually produces a result, or reaction, from the network. For example, if a person walks into a bright room, the nerves in her eyes register the bright light, and send this information to the nervous system. These stimuli pass through the network, resulting in the order for the nerves in the eyes to contract the person's pupils. (Dahr and Stein 1997, p.81)

An artificial neural network consists of a system of *nodes* (neurons) and *weighted connections* (synapses). These are arranged in a number of layers, usually an *input layer*, and *output layer*, and a number of *hidden layers*. The input layer is the layer through which the network can be said to receive the data to be processed, while the output layer displays the result of the network. The hidden layers are where the actual processing of the data takes place. The nodes in the different layers are connected by a series of connections, each assigned a different weight. An example of a simple neural network is illustrated in Figure 6.1. (Dahr and Stein 1997, pp.81-94)

A neural network learns by adjusting the weights of the connections between the nodes. The way in which a neural network learns is called the *learning method*. There are two different kinds of network learning methods: *supervised* and *unsupervised learning*. In supervised learning, the network is provided with set of training data, and the desired outputs for those particular data. Firstly, the network starts out with equal weights on each of the nodes. Then, the data are passed through the network, and the network output is compared to the desired output. If the network output is different from the desired output, the weights of the connections are adjusted, and the simulation is repeated. This process is repeated until the network output matches the desired output. The network then applies what it has learned on any data that are fed into the network in the future.

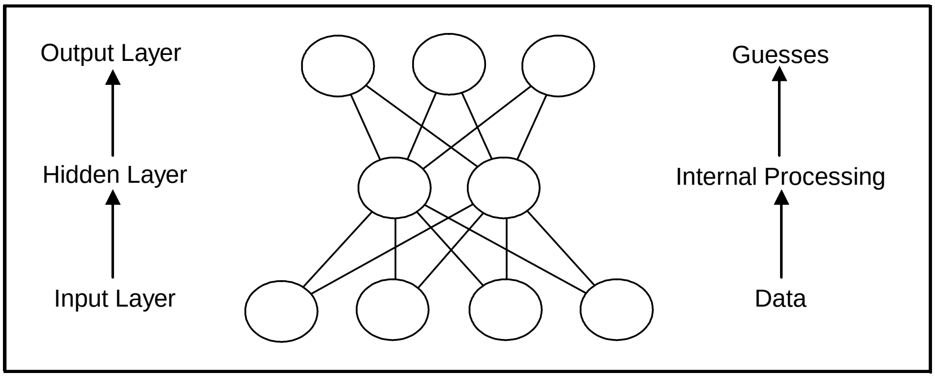


Figure 6.1. A simple neural network (Dahr and Stein 1997, p.83).

The network discussed above is called a *back propagation* network, which is an example of the supervised learning method. The name, back propagation, refers to the fact that the errors between the desired output and the network output are sent back, or *propagated* back, until the desired output is achieved (Dahr and Stein 1997, p.84).

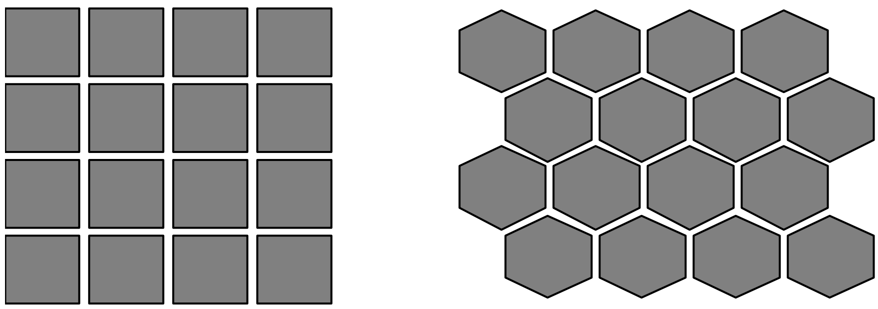
The second kind of learning method and also the method used in this thesis, unsupervised learning, uses a different approach. The network takes the data and clusters them depending upon patterns that it recognizes in the data. Unsupervised networks are usually two layer networks, i.e. with one input layer and one output layer, in which every input node is fully connected to each output node. The learning process, also called *competitive learning*, is characterized by a competition among the neurons. Unlike in supervised learning, in which several nodes can fire at the same time, in unsupervised learning the output neurons all compete to be the single neuron that fires, i.e. the winning neuron (Haykin 1999, p.58). Thus, the neurons in the output layer compete for each row of data presented to the network. Through the learning process, the neurons in the output layer become selectively tuned to specific input patterns, forming groups of similar data. *Self-organizing maps* are an example of applications that use the unsupervised learning method. (Haykin 1999, pp.443-446)

6.3 Self-Organizing Maps

Self-organizing maps (SOMs) are neural networks that use the unsupervised learning method, i.e. the network is presented with input data, and is then allowed to organize itself, depending upon patterns that it recognizes within the input data. Self-organizing maps are two-layer neural networks, i.e. consisting of an input layer and an output layer. The result of the self-organizing process is a "topographic map of the input patterns in which the spatial locations (i.e., coordinates) of the neurons in the lattice are indicative of intrinsic statistical features contained in the input patterns" (Haykin 1999, p.443). The SOM, therefore, essentially performs visual clustering.

6.3.1. The SOM Algorithm

Before the SOM algorithm is initiated, the map is randomly initialized. First, an array of nodes is created. This array can have one or more dimensions, but the most commonly used is the *two-dimensional array*. The two most common forms of lattice are *rectangular* and *hexagonal*, which are also the types used in the SOM_PAK software (see Section 8.1.2). These are illustrated in Figure 6.2. The figure represents rectangular and hexagonal 4×4 lattices, i.e. 16 nodes. In the rectangular lattice, a node has four immediate neighbors, which it interacts with. In the hexagonal lattice, it has six. The hexagonal lattice type is commonly considered better for visualization than the rectangular lattice type. The lattice can also be *irregular*, but this is less commonly used. (Kohonen 1997, p.86)



(a) (b)

Figure 6.2. (a) Rectangular lattice (size 4 x 4) and (b) Hexagonal lattice (size 4 x 4).

Each node i has an associated parametric reference vector m_i. The input data vectors, x, are mapped onto the array. Once this random initialization has been completed, the SOM algorithm is initiated.

The SOM algorithm operates in two steps, which are initiated for each sample in the data set (Kangas 1994, p.15).

Step 1: The input data vector x is compared to the reference vectors m_i, and the best match m_c is located.

Step 2: The nodes within the neighborhood h_{ci} of c are "tuned" to the input data vector x.

These steps are repeated for the entire dataset, until a stopping criterion is reached, which can be either a predetermined amount of trials, or that the changes are small enough.

In step 1, the best matching node to the input vector is found. The best matching node is determined using some form of distance function, for example, the *smallest Euclidian distance* function, defined as $\|x - m_i\|$. The best match, m_c, is found by using the formula in Equation 6.1 (Kohonen 1997, p.86):

Equation 6.1

$$\|x - m_c\| = \min_i \{\|x - m_i\|\}.$$

Once the best match, or winner, is found, step 2 is initiated. This is the "learning step", in which the network surrounding node c is adjusted towards the input data vector. Nodes within a specified geometric distance, h_{ci}, will activate each other, and learn something from the same input vector x. This will have a smoothing effect on the reference vectors in this neighborhood. The number of nodes affected depends upon the type of lattice and the neighborhood function. This learning process can be defined as (Kohonen 1997, p.87):

Equation 6.2

$$m_i(t+1) = m_i(t) + h_{ci}(t)[x(t) - m_i(t)],$$

where t = 0,1,2,... is an integer, the discrete-time coordinate. The function $h_{ci}(t)$ is the neighborhood of the winning neuron c, and acts as the so-called *neighborhood function*, a smoothing kernel defined over the lattice points. The function $h_{ci}(t)$ can be defined in two ways. It can be defined as a neighborhood set of arrays around node c, denoted N_c, whereby $h_{ci}(t) = \alpha(t)$ if $i \in N_c$, and $h_{ci}(t) = 0$ if $i \notin N_c$. Here $\alpha(t)$ is defined as a *learning rate factor* (between 0 and 1). N_c can also be defined as a function of time, $N_c(t)$.

The function $h_{ci}(t)$ can also be defined as a *Gaussian function*, denoted:

Equation 6.3

$$h_{ci} = \alpha(t) \cdot \exp\left(-\frac{\|r_c - r_i\|^2}{2\sigma^2(t)} \right),$$

where $\alpha(t)$ is again a learning rate factor, and the parameter $\sigma(t)$ defines the width of the kernel, or radius of $N_c(t)$.

For small networks, the choice of process parameters is not very important, and the simpler neighborhood-set function for $h_{ci}(t)$ is, therefore, preferable. (Kohonen 1997, p.88)

However, when using the simpler neighborhood set function $N_c(t)$, the choice of neighborhood radius is very important: if the initial neighborhood is too small, the map will not be ordered globally. This will result in mosaic-like patterns on the map, between which no ordering is noticeable. Therefore, it is better to start out with a large neighborhood, and let it shrink over time. The initial radius can even be larger than half of the diameter of the map. (Kohonen 1997, p.88)

The training process is illustrated in Figure 6.3. The figure shows a part of a hexagonal SOM. Firstly, the reference vectors are mapped randomly onto a two-dimensional, hexagonal lattice. This is illustrated in Figure 6.3 (a) by the reference vectors, illustrated by arrows in the nodes, pointing in random directions. In Figure 6.3 (a) the closest match to the input data vector *x* has been found in node *c* (Step 1). The nodes within the neighborhood h_{ci} learn from node *c* (Step 2). The size of the neighborhood h_{ci} is determined by the parameter $N_c(t)$, which is the neighborhood radius. The reference vectors within the neighborhood h_{ci} tune to, or learn from, the input data vector *x*. How much the vectors learn depends upon the learning rate factor $\alpha(t)$. In Figure 6.3 (b), the final, fully trained network is displayed. In a fully trained network, a number of groups should have emerged, with the reference vectors between the groups “flowing” smoothly into the different groups. If the neighborhood h_{ci} were to be too small, small groups of trained reference vectors would emerge, with largely untrained vectors in between, i.e. the arrows would not flow uniformly into each other. Figure 6.3 (b) is an example of a well-trained network.

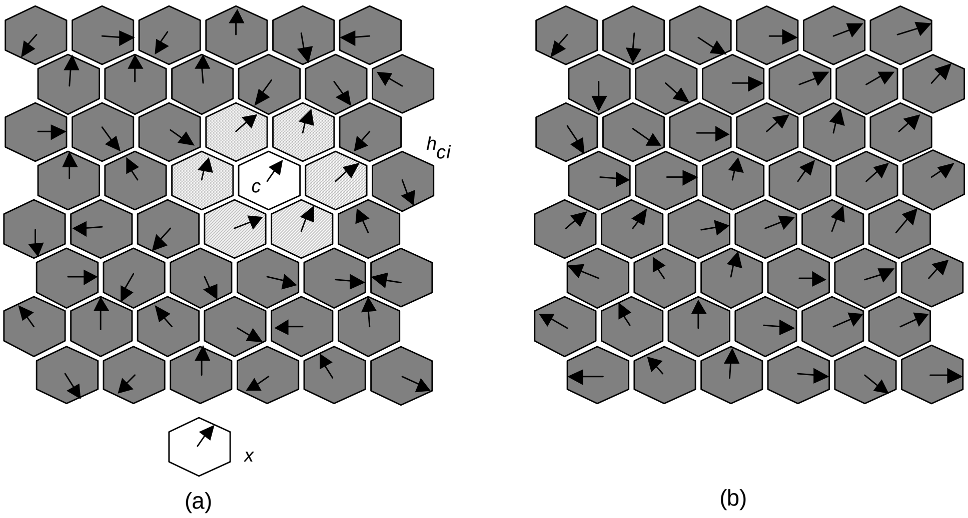


Figure 6.3. (a) A randomly initialized network after one learning step and (b) a fully trained network (Kohonen 1997, p.92).

There are a number of software packages in which the SOM algorithm is implemented. Some of the packages are commercial, while others are available for scientific use. Programs available for scientific use include SOM_PAK (Kohonen et al. 1996), a UNIX or DOS application, and a toolbox for Matlab 5, the SOM Toolbox (Vesanto et al. 2000). Both of these programs have been developed by the Neural Networks Research Center at the Helsinki University of Technology. Commercial products include, for example, Nenet v1.1 (see Section 8.1.2), NeuroSolutions 4.2 (www.nd.com), SOMine (www.euadaptics.com), Visipoint (www.visipoint.fi), and eSOM (www.ellipse.fi).

6.3.2. The Average Quantization Error

The most common way of measuring the quality of a map is by calculating the *average quantization error*, *E*. The average quantization error represents the average distance between the best matching units and the sample data vectors. The average quantization error can be calculated using the formula:

Equation 6.4

$$E = \frac{1}{N}\sum_{i=1}^{N} \min_{c}\left\{\left\| x_i - m_c \right\|\right\},$$

where N is the total number of samples, x_i is the input data vector, and m_c is the best matching reference vector.

6.3.3. Visualizing the Maps

Once a self-organizing map has been created, it must be visualized in order for it to be interpreted. *Unified distance matrix*, or *U-matrix* (Ultsch 1993), is the most common way of visualizing self-organizing maps. U-matrix maps are created by calculating the average of the distances of a reference vector in a node to that of its neighboring reference vectors. This average is placed at the appropriate coordinate on the matrix. The shape of the matrix is dependent upon the neighborhood topology, i.e. rectangular or hexagonal. The result is basically a three-dimensional landscape, with peaks and valleys. Peaks, or great distances between reference vectors, are displayed as dark areas on the map, while valleys, or light areas, represent short distances (Ultsch 1993). Therefore, clusters with lighter shaded areas between them are closer to each other than clusters with darker shades between them. This effectively allows us to locate similar units on the map, and to identify groups of similar units. This process can be referred to as *clustering via visualization* (Flexer 2001), i.e. the SOM is used to cluster the data, and the clusters are subjectively isolated by studying the visualization of the topology.

There are two ways to display U-matrix maps: in *grayscale* (Figure 6.4 (a)) or in *color* (Figure 6.4 (b)). The grayscale U-matrix is used in the SOM_PAK software, while the color U-matrix is used, for example, in the Nenet software. SOM_PAK and Nenet are presented in Section 8.1.2.

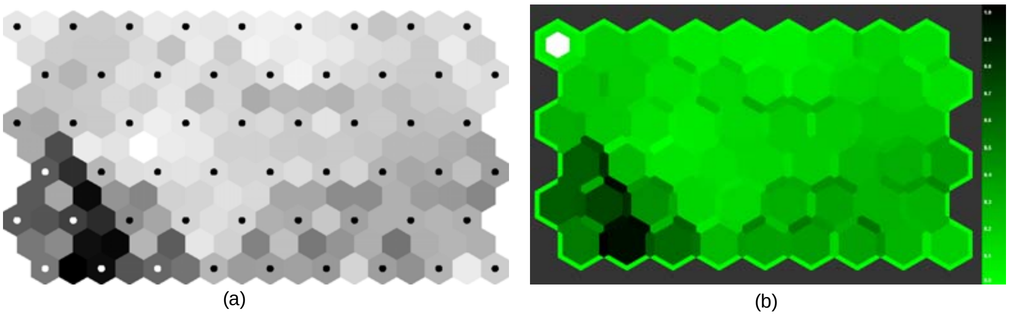
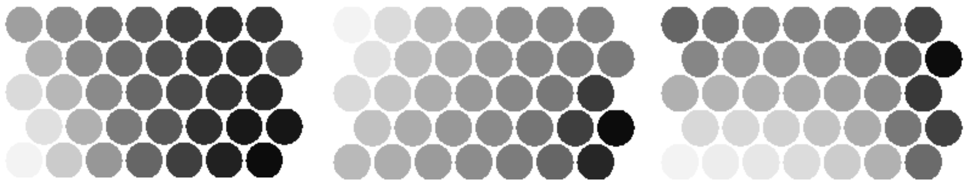


Figure 6.4. The same U-matrix in (a) grayscale (SOM_PAK) and (b) color (Nenet v1.1).

6.3.4. Feature Planes

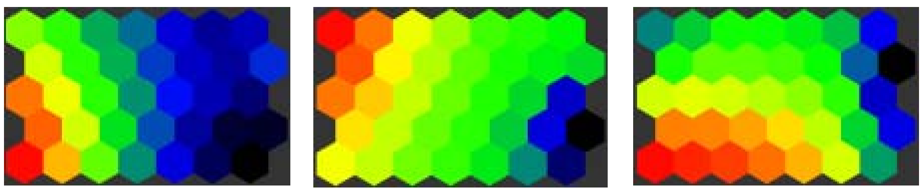
In addition to U-matrix maps, single vector-level maps, called *feature planes*, can also be created. These maps display the distribution of individual columns of data, in this case the values of individual financial ratios. These maps are used to identify the characteristics of the clusters on the maps. Three examples of feature planes are illustrated below (Figure 6.5 (a), (b), and (c)). Light colors identify high values, and dark colors identify low values.



(a) (b) (c)

Figure 6.5. (a) Operating Margin, (b) Return on Equity, and (c) Equity to Capital feature planes.

The feature planes can also be displayed in color, making them easier to interpret. The same feature planes are presented in color in Figure 6.6. On these feature planes, "warm" colors identify high values, and "cold" colors identify low values. This is the method for displaying feature planes in Nenet v1.1.



(a) (b) (c)

Figure 6.6. (a) Operating Margin, (b) Return on Equity, and (c) Equity to Capital feature planes.

6.3.5. Trajectories

When viewing data as a time-series or process, it is interesting and important to visualize the changes in state. This can be done using trajectories (Figure 6.7) as is proposed by Alhoniemi et al. (1999) and Simula et al. (1999b). Trajectories have been used in many SOM applications, for example, in process state visualization (Tryba and Goser 1991; Himberg et al. 2001). It has also been by Oja and Kiviluoto (1999) to illustrate the financial state of companies in bankruptcy prediction. As in Oja and Kiviluoto (1999), in this case trajectories are used to illustrate the financial situation of a company at a point in time, as well as the following state changes.

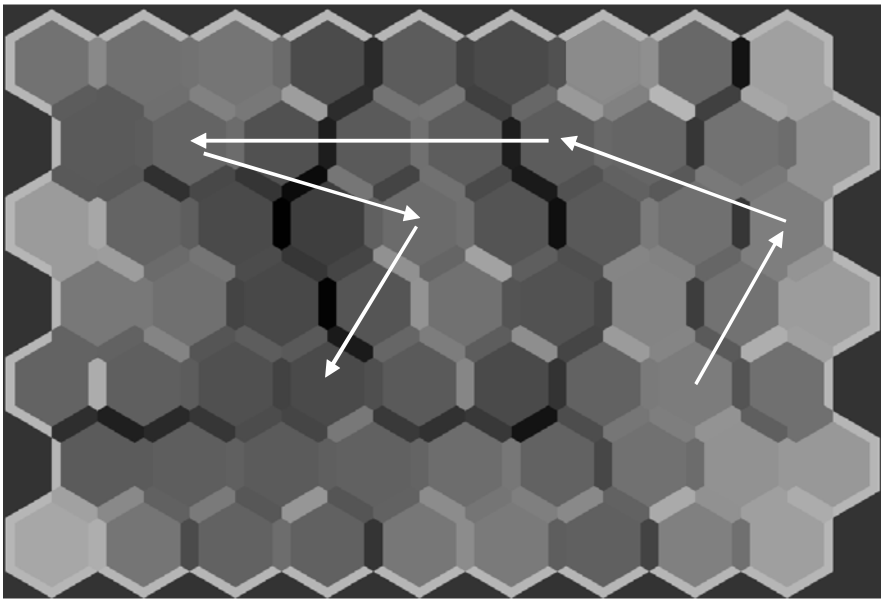


Figure 6.7. Example of the use of trajectories.

7 THE SELECTED SET OF COMPANIES

This section describes and justifies the choice of companies used in this experiment. The data collection process is described, and rules for dealing with missing data are also presented. Finally, a list of the companies included is shown.

7.1 Choosing the Companies

The companies to be used in this experiment were chosen based on *Pulp and Paper International's* annual Top 150 report (Rhiannon et al. 2001). Every year, in their September issue, *Pulp and Paper International*, a leading paper industry publication, publishes a report in which they rank the top 150 pulp and paper producing companies in the world according to net sales.

One problem that must be dealt with is which companies to include, as many of them have operations outside pulp, paper, and converting. However, as it is impossible to obtain complete financial data for only pulp, paper, or converting operations, the consolidated financial statements have been used to obtain the required data. This, of course, results in other operations being included in the experiment, but there is no practical way of preventing this. Besides, most of the companies involved in the experiment have other operations that in some way relate to pulp and paper. For example, while Ahlstrom Paper Group only accounted for 51% of Ahlstrom's total operations in 1999, two of the other groups, Ahlstrom Machinery Group and Ahlstrom Pumps, each specialize in machinery for the pulp and paper industry, and the Åkerlund & Rausing Group specializes in consumer packaging products. Since such a large part of their operations were related to pulp and paper, the consolidated statements have been used to include the company in this experiment. However, there was an exception to this rule: Procter & Gamble is listed as number three on the Top 5 list, but pulp, paper, and converting operations only account for 30% of the company's consolidated results. Since pulp and paper operations accounted for less than 50% of operations, Procter & Gamble was left out.

7.2 Collecting the Data

As the original intent of the experiment was to use publicly available sources for financial information, the annual reports were collected primarily from companies' own homepages. However, the companies rarely provided enough information on their own pages. There was also a very large difference between the homepages of companies in different countries; most US and Canadian companies provided far more information on their homepages than European (besides Scandinavian companies) or Japanese companies did. After collecting all of the information available on companies' homepages, the only companies with adequate financial data were from the USA, Canada, Sweden, or Finland. Surprisingly many companies provided no online information at all, or only abbreviated versions of their reports.

Therefore, three online databases were used to complete the information: *EDGAR* (http://www.sec.gov/) for US reports, *SEDAR* (http://www.sedar.com/) for Canadian reports, and *Japan Financials* (http://www.japanfinancials.com/) for Japanese reports.

The EDGAR (Electronic Data Gathering, Analysis and Retrieval system) database is an online database for various electronic financial reports submitted to the *US Securities and Exchange Commission* (SEC), as is required by law for all companies noted on the New York Stock Exchange (NYSE). These reports include copies of the companies' annual reports, known as *10-K filings*.

SEDAR, (System for Electronic Document Analysis and Retrieval), is the Canadian equivalent of EDGAR. The service is provided by the *Canadian Depository for Securities*. SEDAR only contains annual reports, which are stored in Portable Document Format (PDF).

Japan Financials is a service provided by *Eagle Enterprises Ltd*. Japan Financials provides English translations of the balance sheets and income statements of publicly traded companies in Japan. The financial statements are found in the *yuka shoken hokokusho,* or annual report, filed with the Japanese equivalent of the US SEC. In 2001, Japan Financials.com started charging for its services, making annual reports for 2000 unavailable through this source.

A major problem was that no database containing information on European companies could be found. Therefore, the data were completed with any physical reports that could be obtained from the companies themselves.

7.3 Missing and Incomplete Data

In many cases, the information available was not sufficient for the experiment. This could be due to the fact that a company did not provide enough information in its annual report, or that the annual report for a particular year (usually 1995) was not available. In cases where only some information was missing, Instead of just discarding the entire company, estimates have been used. This was typical of data for 1994, which was only used to calculate averages to be used in three ratios. When the annual reports for different years were compared, it was in many cases possible to observe a trend in the development of certain values. In many cases it was towards growth, but in some cases it was even a decline. However, since it was in most cases possible to observe a trend, an averaging formula could be used to calculate estimates for missing data.

In order to obtain estimates, the available data were used to calculate the annual change. This is shown in Equation 7.1.

Equation 7.1

$$d = \frac{x_{i+1} - x_i}{x_i} .$$

In Equation 7.1, x_i stands for the existing value for one year and x_{i+1} for the existing value for the following year. The result, *d*, is the change from the one year to the next. Wherever an existing value for a particular year differed significantly from other years, it was left out of the equation. The annual change was then averaged over the five-year data range. The formula used can be found in Equation 7.2.

Equation 7.2

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i .$$

This formula is simply the sample mean of the calculated annual changes. Finally, the missing value was calculated by subtracting the calculated annual change, $\bar{d}$, from the earliest value which was available, as seen below.

Equation 7.3

$$x_{i-1} = x_i - (x_i \times \bar{d}) .$$

Here, x_i is the value for the first year for which data is available, and x_{i-1} is the estimated value for the previous year. The calculated value was then compared to data for other years, to see if it seemed feasible. The same procedure was used wherever small amounts of data were missing. This allowed for companies with some missing data to be included in the experiment. This, of course, led to some data not being completely accurate, but this simplification was necessary in order to be able to include substantial amounts of data for the year 1995.

7.4 Companies Included

The companies included in the experiment are presented in Table 7.1 and Table 7.2. In the end, 91 companies and 7 regional averages were included. It was only possible to include a small group of European companies, since annual reports for these companies were very hard to obtain. The greatest disappointment was that not a single non-Japanese company from Asia, primarily China, was possible to include. In addition, German companies had to be left out, which was also a loss. However, the selected companies represent a good overall selection considering that the four largest paper producers in the world are USA, Japan, China, and Canada (Finnish Forest Industries Federation 2000). Also, some very important companies had to be left out for various reasons. As was mentioned earlier (Section 7.1), Procter & Gamble had to be left out. In addition, Nippon Unipac Holding was formed during 2001, and was still regarded as two separate companies (Nippon Paper Industries and Daishowa Paper) during this experiment.

A number of the companies included have been included as single companies, although technically, they were still separate companies during the period in question. This was done in order not to have two companies for only a single year, which then merge during the following year, since this reduces overall comparability. This simplification was made possible by the fact that companies merging present consolidated annual reports dating back to before the merger. Examples of companies treated in this way include United Paper Mills and Kymmene Oy who merged in 1996 to form UPM-Kymmene, and Stora and Enso, who merged in 1998 to form Stora Enso.

Finland			14	Trebruk	98-01
1	Average		Norway		
2	Ahlström	95-01	15	Average	
3	M-Real	95-01	16	Peterson Group	95-01
4	Stora Enso OY (Enso Oy 95-96)	97-01	17	Norske Skog A.S.	95-01
5	UPM-Kymmene OY	95-01	USA		
Sweden			18	Average	
6	Average		19	Boise Cascade	95-01
7	Sveaskog (AssiDomän 95-00)	95-01	20	Bowater	95-01
8	Korsnäs	95-01	21	Buckeye Techologies	95-01
9	MoDo AB	95-01	22	Caraustar Industries	95-01
10	Munksjö AB	95-01	23	Champion International	95-99
11	Rottneros AB	95-01	24	Consolidated Papers	95-99
12	SCA AB	95-01	25	Crown Vantage	95-99
13	Södra AB	95-01	26	FiberMark	95-01

Table 7.1. Companies used in this experiment.

27	Fort James	95-99	64	Tembec Inc.	95-01
28	Gaylord Container Corp	95-01	65	West Fraser Timber	95-01
29	Georgia-Pacific Corp	95-01	Japan		
30	Greif Bros	95-01	66	Average	
31	International Paper	95-01	67	Daio Paper	95-99
32	Jefferson-Smurfit Corp.	95-01	68	Daishowa Paper Manuf	95-99
33	Kimberly-Clark	95-01	69	Chuetsu Paper	95-99
34	Longview Fiber Corp.	95-01	70	Hokuetsu Paper Mills	95-99
35	Mead	95-00	71	Japan Paperboard Industries	95-99
36	Packaging Corp	99-01	72	Kishu Paper	97-99
37	P.H. Glatfelter	95-01	73	Mitsubishi Paper	95-99
38	Pope & Talbot	95-01	74	Nippon Kakoh Seishi	95-99
39	Potlatch Corp.	95-01	75	Nippon Unipac (Nippon Paper - 00)	95-01
40	Rayonier	95-01	76	Oji Paper	95-01
41	Riverwood Holding	95-01	77	Rengo	95-99
42	Rock-Tenn Company	95-01	78	Settsu	95-98
43	Schweitzer-Mauduit Intl.	95-01	79	Tokai Pulp & Paper	95-99
44	Sonoco Products	95-01	Europe and others		
45	Stone Container	95-97	80	Average	
46	Temple-Inland	95-01	81	Cartiere Burgo (ITA)	95-01
47	Union Camp.	95-98	82	David S. Smith Holdings (UK)	98-01
48	Wausau-Mosinee Paper	95-01	83	ENCE Group (Spain)	96-01
49	MeadWestvaco (Westvaco 95-00)	95-01	84	Execompta Clairefontaine (FRA)	97-01
50	Weyerhaeuser	95-01	85	Frantschach (AUT)	95-99
51	Willamette Industries	95-01	86	Gascogne (FRA)	98-01
Canada			87	Kappa (NLD)	98-01
52	Average		88	Industrieholding Cham (SUI)	95-01
53	Abitibi Consolidated	95-01	89	Inveresk (UK)	95-01
54	Alliance	95-00	90	Mayr-Melnhof (AUT)	95-01
55	Canfor	95-01	91	Mercer International (SUI)	95-01
56	Cascades Inc.	95-01	92	Reno de Medici (ITA)	95-01
57	Crestbrook Forest Ind.Ltd.	95-97	93	Aracruz Celulose (BRA)	96-01
58	Doman Industries	95-01	94	Amcor (AUS)	95-00
59	Domtar Inc.	95-01	95	Bahia Sul Selulose (BRA)	98-01
60	Donohue	95-99	96	Empresas CMPC (CHL)	98-01
61	MacMillan Bloedel	95-98	97	Fletcher Challenge Group (NZE)	95-99
62	Millar Western Forest Products	98-01	98	Sappi (ZAF)	98-01
63	Nexfor	95-01			

Table 7.2. Companies used in this experiment (continued).

8 THE BENCHMARKING MODEL

This chapter describes the final model evaluated in this thesis. The first section deals with the training of the map. The second section provides an analysis of the map, while the third section shows a sample benchmarking of the Top 5 pulp and paper companies. Multilevel analysis and combining quantitative and qualitative analysis are also discussed in this chapter.

8.1 Training the Network

This section describes the process of training the network. Firstly, different methods for preprocessing the data are discussed. Then, the software packages used are introduced. The selection of parameters during the training process is then presented. Finally, the construction of the final map is explained.

8.1.1. Data Preprocessing

In many cases, it is necessary to rescale the data in order to ease the network's learning process. If the variation in one value is significantly higher than in another, the network will expend its learning time on the first value, possibly ignoring the later value. In addition, it has been shown in several studies that financial ratios generally do not follow a normal distribution, and that outliers are common (Lev 1974, pp.61-62; Foster 1978, pp.170-179; Salmi and Martikainen 1994). In this experiment, the differences were large; for example, Equity to Capital ratios varied between -38.37 and 112, while, for example, Receivables Turnover ratios varied between 0.14 and 20.81. The differences in scale between the ratios are illustrated in Table 8.1. These differences mean that the network will have a hard time settling on a good solution, i.e. each time a new learning step is run, the map continues to change greatly.

	Operating Margin	ROE	ROTA	Equity to Capital	Quick Ratio	Interest Coverage	Receivables turnover
Average	9.20	-12.69	7.62	37.42	1.61	4.88	7.33
Variance	58.81	75003.13	49.36	304.73	7.39	157.13	9.30

Table 8.1. Averages and variances for the different ratios included.

The table shows that the variance among Equity to Capital and ROE ratios is extremely large. Especially the high variance of the ROE ratio is problematic. As there was no way in which to force the ratios to produce similarly scaled values, a suitable rescaling method had to be used, i.e. the data had to be *preprocessed* according to some method. Shanker et al. (1996) performed an experiment comparing the performance of a neural network on unstandardized data as opposed to data standardized using linear transformation and normalization. The authors conclude that the networks using standardized data outperformed the network using unstandardized data in all test cases, especially when using small datasets. In addition, normalization was found to be much better than linear transformation.

The choice of preprocessing method is a much-discussed subject in literature concerning neural networks. Sarle (2001) suggests that the most suitable form of standardization centers the input values around zero, instead of, for example, within the interval [0,1]. This would imply the use of, for example, *normalization by standard deviation,* also proposed by Shanker et al. (1996). In addition to normalization by standard deviation, Kohonen (1997, p.121) also suggests the use of heuristically justifiable rescaling. Another method, suggested by Guiver and Klimasauskas (1991), and used, for example, by Back et al. (1998b), Back et al. (2001), and Kiviluoto (1998), is *histogram equalization*. Histogram equalization is a method for mapping rare events to a small part of the data range, and spreading out frequent events. This way the network is better able to discriminate between rare and frequent events.

Normally, the values should be scaled according to their relative importance. However, according to Kaski and Kohonen (1996), this is not necessary when no differences in importance can be assumed. As this is also the case in this experiment, no form of scaling according to importance has been used. Instead, the relative importance of the different categories of ratios has been set through the balance of ratios (three in profitability, two in solvency, one in liquidity, and one in efficiency).

During the course of this experiment, several different preprocessing methods were experimented with. The first attempt was to normalize the variables one by one, using the standard deviation. However, this did not lead to satisfactory maps, since the maps were very difficult to interpret, and the average quantization error was very high. The same was then attempted using the variance, but although the average quantization error was much lower, the maps were still difficult to interpret. In the original experiment (Paper 2), the data were normalized according to the variance over the entire dataset. Although this method gave acceptable maps, modifications to the data, in the form of removal of extreme values, were required. Although there are methods for determining outliers and extreme values, such as modified box-plots, a method that does not require additional modifications to the data would be preferable. Histogram equalization has this advantage over normalization. Also, using the variance over the entire dataset, some ratios, primarily the quick ratio, received very little attention during the training cycle. This was due to the fact that the range of this ratio was much smaller than that of the others. The same can be said of turnover efficiency. For this experiment, histogram equalization was therefore chosen for use.

As was mentioned above, histogram equalization is a method for mapping rare events to a small part of the data range, and spreading out frequent events. Histogram equalization is a technique more commonly associated with image processing. Within image processing, it is used to increase contrast in black and white images. In some images, the different scales of gray are not evenly distributed, i.e. some scales are not used at all, creating very little contrast in an image. Histogram equalization can be used to spread out the scales, allowing more effective use of the different shades of gray.

Histogram equalization works by creating a histogram of the frequency of occurrence of values in the dataset. Firstly, a histogram consisting of a specified number of bins is created. Then, each variable in the dataset is assigned to one bin, incrementing the value of that bin by one. Then, a cumulative histogram is created by adding to each bin the value of its preceding bins. Finally, the value of each bin is divided by the total number of observations, thus standardizing the data within the interval of [0,1] (other intervals can also be used).

Histogram equalization can be illustrated using Guiver and Klimasauskas' (1991) example concerning credit cards. In the example, a fictive histogram representing the number of credit cards in a single household is used to illustrate the function of histogram equalization. Table 8.2 shows how many credit cards the average household has, and a histogram of the frequency of each observation. The example shows, for example, that 84 households have 4 credit cards.

Number of Cards	0	1	2	3	4	5
Occurrences	0	10	85	67	84	161

Table 8.2. Number of credit cards in the average household.

After this, a cumulative histogram is created by summing each number of occurrences with the number of occurrences of the previous observations, as in Table 8.3.

Cumulative Hist.	0	10	95	162	246	407

Table 8.3. A cumulative histogram.

Finally, Equation 8.1 is applied.

Equation 8.1

$$C_i \div T,$$

where C_i is the *i*th value in the cumulative histogram, and T is the cumulative total in the histogram. In this case, the resulting equalized data can be found in Table 8.4.

Equalized data	0.000	0.025	0.233	0.398	0.604	1.000

Table 8.4. Histogram equalized data in the credit card example.

As can be seen in Table 8.4, the data is now equalized within the interval [0,1].

In this study, histogram equalization has been used to preprocess the data. The choice of preprocessing method is heavily dependent upon the overall intention of the project. In this case, this provided me with clear clusters to analyze. When using normalization according to the standard deviation or variance, this resulted in maps that were flat except for some regions on the extreme ends of the maps. In order to use these methods, it would have been necessary to remove peaks in the values, as was done in Paper 2. However, histogram equalization can automatically cope with problems like this, and the data therefore needs no other preprocessing, apart from normal cleaning (removal of errors, incorrect formats, and duplicate values).

One must remember that cluster analysis is only one possible application of self-organizing maps. While we in this case are interested in grouping companies according to similarities, we might just want to map them on a large map to compare individual characteristics by studying the feature planes. In such cases, even if the map may appear flat, each neuron is better tuned to a specific combination of characteristics. Therefore, the results on an individual feature plane can be said to be more accurate than in this case, since there is no forcing of data into clusters. If this had been the purpose in this study, normalization according to standard deviation or variance would have been quite suitable alternatives.

8.1.2. The Program Packages Used

The software package used for training the self-organizing maps in this thesis is called *The Self-Organizing Map Program Package*, or SOM_PAK. SOM_PAK is a program package developed by the *Neural Networks Research Center* (NNRC) at the Helsinki University of Technology. The program package is free for non-commercial use. The version of SOM_PAK used in this experiment is version 3.1. SOM_PAK consists of a number of separate programs used for the training process. Each step of the process can be run using separate programs, but SOM_PAK also includes a program, vfind.exe, which runs the entire training process at once, by requesting all of the parameters before starting the training cycle.

Although a program for visualizing the maps is included in the SOM_PAK package, a separate program has been used for this purpose. This program is called Nenet v1.1, and is available as a limited demo at http://koti.mbnet.fi/~phodju/nenet/Nenet/General.html. The advantage with using Nenet as opposed to SOM_PAK for visualization of maps is that Nenet produces maps in shades of green, instead of grayscale. This, in our opinion, makes them easier to read. Therefore, all maps displayed in this thesis will be displayed using Nenet.

Finally, Viscovery SOMine 4.0 has been used to automatically identify the clusters on the map using two-level clustering. In this case, Ward's method has been used (see Section 8.1.4).

8.1.3. Selecting the Parameters

Although the optimal parameters are different in each case, there are a number of recommendations for parameters used in the training process. These are actually more like starting points, from which to work out the optimal parameters for the experiment in particular. When training small maps (less than a few hundred nodes), the selection of parameters does not greatly influence the outcome of the training process (Kohonen 1997, p.88). There are, however, a number of recommendations for training maps, which should be noted. These recommendations will be discussed below.

The network topology refers to the shape of the lattice, i.e. rectangular or hexagonal. The topology should in this case be hexagonal, since hexagonal lattices are better for visualization purposes (Kohonen 1997, p.86).

Network size, or the dimensions of the map, is important for visualization purposes. If the map is too small, differences between units are hard to identify. Smaller changes from year to year are also difficult to illustrate. However, a small map is best for cluster identification purposes. On the other hand, if the map is too large, the clusters do not appear, and the map seems "flat". The size of the map is thus determined by the purpose of the experiment (Deboeck 1998, p.208). Another thing to remember is that the map dimensions should be rectangular instead of square. This is because the reference vectors must be oriented along with $p(x)$ in order for the network to stabilize during the learning process (Kohonen et al. 1996). A commonly used principle is that $p(x)$ should be roughly 1.3 times the length of *p(y)*, where *p(x)* is the length of the x-axis and *p(y)* is the length of the y-axis.

The statistical accuracy of the mapping depends upon the number of steps in the final learning phase. This phase therefore has to be relatively large. A good rule of thumb is that in order to achieve good statistical accuracy, the amount of steps in the final phase must be at least 500 times the amount of nodes in the network (Kohonen 1997, p.88). It is common practice for the initial training phase to have at least 10 percent of the amount of steps used in the final phase.

The learning rate factor, or $\alpha(t)$, should start out as fairly large in the first phase, but should be very low in the final phase. A commonly used starting point is 0.5 for the first phase, and 0.05 in the final phase.

The selection of the network neighborhood size, $N_C(t)$, is possibly the most important parameter. If the selected neighborhood size is too small, the network will not be ordered globally. This will result in various mosaic-like patterns, with unordered data in between. Therefore, the initial network radius should be rather large, preferably larger than half the network diameter (Kohonen 1997, p.88). Generally, the final network radius should be about 10% of the radius used in the first part.

8.1.4. Constructing the Maps

Several hundred maps were trained during the course of the experiment. The first maps were trained using parameters selected according to the guidelines presented in Section 8.1.3. The best maps, rated according to quantization error and ease of readability, were then selected and used as a basis when training further maps. A smaller map could have been used if a separate map had been trained for each year, but the intention was to use the same map for the entire dataset. A 9 x 7 sized map seemed large enough to incorporate the data for each year included in the test. The 9 x 7 lattice also conforms to the recommendation that $p(x) = 1.3 \times p(y)$.

The number of steps used in the final phase was generated directly from the recommendations provided above. Therefore, the initial phase includes 3,150 steps and the final phase 31,500 steps. The learning rate factor was set to 0.5 in the first phase and 0.06 in the second, very near the recommended starting point. The neighborhood radius was set to 11 for the first phase and 1 for the second. The initial radius was very large compared to the recommendations, but seemed to provide for the overall best maps. Decreasing the radius only resulted in poorer maps.

As Kohonen noted (Kohonen 1997, p.88), the selection of parameters appears to make little difference in the outcome when training small maps. As long as the initial selected parameters remained near the guidelines presented above, the changes in the quantization error were very small, usually as little as 0.001. Some examples of the parameters and outcomes are illustrated in Table 8.5. These are only a fraction of the entire training set, but illustrate well the small differences in results. One must note that maps with differing dimensions or radii can not be compared using the quantization error as a measure of goodness (Kaski and Lagus 1996).

	Map1	Map2	Map3	Map4	Map5	Map6	Map7	Map8
Size	9x7	9x7	9x7	9x7	9x7	9x7	9x7	9x7
Training length 1	3,150	3,150	3,150	3,150	3,150	4,000	3,150	3,150
Training rate 1	0.5	0.6	0.4	0.5	0.5	0.5	0.5	0.5
Radius 1	11	11	11	10	11	11	11	11
Training length 2	31,500	31,500	31,500	31,500	31,500	31,500	40,000	31,500
Training rate 2	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.06
Radius 2	1	1	1	1	2	1	1	1
Error	0.3264	0.3282	0.3330	0.3302	0.3277	0.3282	0.3270	0.3281

Table 8.5. Examples of trained 9x7 maps.

Table 8.5 shows that the changes are very small irrespective of the parameters used. The map that was finally chosen was map 8. It is notable that this map was trained using parameters generated directly from the recommendations in Section 8.1.3, with the exception of the training rate in part 2. Although there are several maps with lower quantization errors, the differences are small, and map 8 was considered to be the most easily interpretable of the trained maps. The appearance of the maps was monitored throughout the experiment, but very small differences in the resulting maps surfaced. Although the maps might look slightly different, the same clusters containing approximately the same companies were found in the same positions relative to each other. While the “good” end of one map might have been found on the opposite side of another map, the same clusters could still be seen to emerge. This shows the random initialization process of the self-organizing map, but also proves that the results from one map to another are consistent.

Vesanto and Alhoniemi (2000) and Siponen et al. (2001) explore automatic clustering of the SOM, and the use of hierarchical clustering methods is proposed in order to determine the clusters in the data. Viscovery SOMine (http://ww.eudaptics.com) allows the user to perform cluster identification using a number of methods, including Ward’s method, an agglomerative hierarchical clustering technique. It also suggests the number of “natural” clusters in the data.

In this case, these capabilities have been used to cluster the map. Ward's clustering has been used for this purpose. This increases the objectivity of the cluster analysis.

As was mentioned above, the maps were trained by using the entire dataset, from 1995-01, to create a single map. Separate maps for each year were then produced, by labeling only the companies for the year in question. This allowed me to use the same U-matrix map for each year. Thus, the clusters were always located in the same place on the map each year, making it much easier to identify the best performing companies.

8.2 Identifying the Clusters on the Map

By studying the underlying feature planes (Figure 8.3), a number of clusters of companies, and the characteristics of these clusters, can be identified. Figure 8.2 summarizes the most important areas of the map. The positions of the groups are presented in Figure 8.1. The groups are presented below.

Group A can be seen as the best performing group. The group has very high values in all profitability ratios. The group also shows high or very high values in all other ratios, except for Quick Ratio, which varies between average and very high.

Group B is a well above average group, very similar to Group A. The values of all ratios are nearly as high as in Group A, except for solvency, which is somewhat lower, and liquidity, which is very low.

Group C is a slightly above average group. The group has high values in profitability, but low solvency. Liquidity is average.

Group D is an average group. Profitability in this group is average, although there are some high values in operating margin. Solvency is average, and efficiency is high.

Group E is characterized by very high solvency and liquidity values. Profitability is low to average.

Group F is a below average group. Profitability and solvency are low to average, and liquidity is very low. Efficiency can reach high values.

Group G is one of the two really poor groups. The group is characterized by very low values in all ratios, although profitability is slightly better than in Group H.

Group H is the other poor group. Group H contains the poorest companies in terms of profitability and solvency. However, liquidity and especially efficiency can be high.

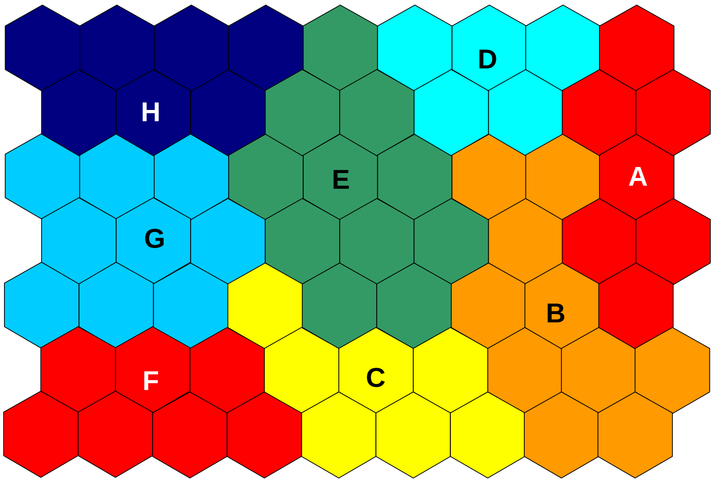


Figure 8.1. Clusters identified on the map.

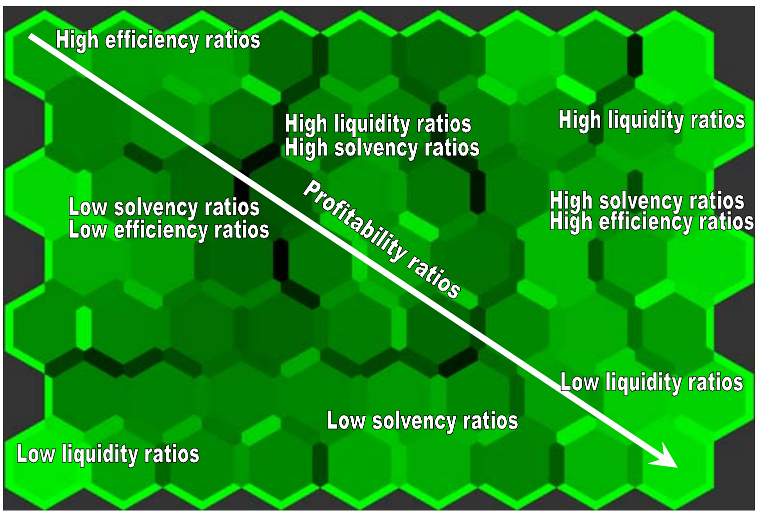


Figure 8.2. Areas of importance on the map.

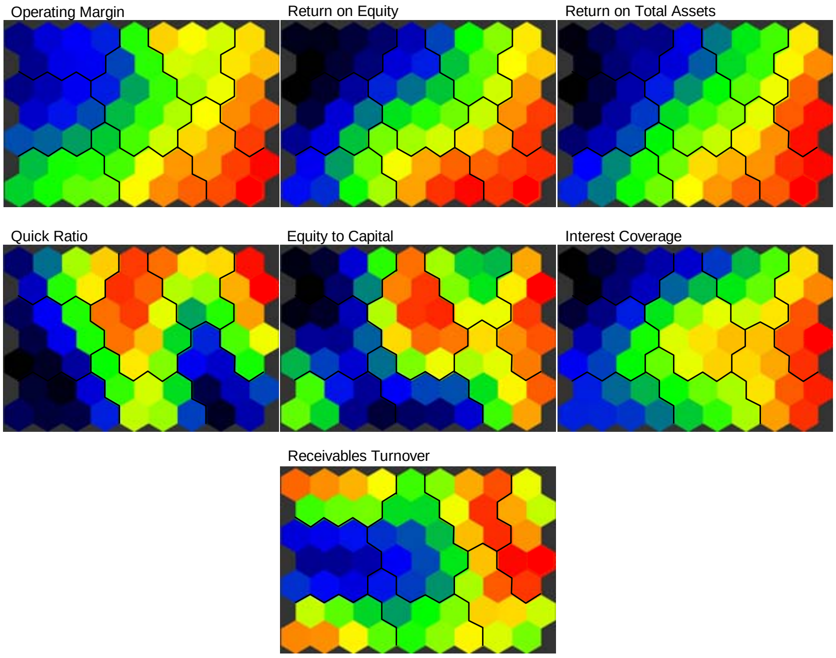


Figure 8.3. The feature planes of the final map.

	A	B	C	D	E	F	G	H
Operating Margin	VH	H-VH	M-VH	M-H	L-M	M	L	L
ROE	VH	H-VH	M-VH	M	L-M	L-M	VL-M	VL
ROTA	VH	H-VH	M-VH	M	L-M	L-M	VL-L	VL
Equity to Capital	VH	M-VH	L	M	H-VH	VL-M	VL-L	VL-L
Quick Ratio	M-VH	VL-L	M	M-VH	H-VH	VL	VL-M	VL-H
Interest Coverage	VH	VH	M	M	L-H	L	VL-M	VL
Receivables Turnover	H-VH	M-VH	M	H-VH	L-M	M-H	L	M-VH

Table 8.6. Cluster descriptions (VL = very low, L = low, M = medium, H = high, and VH = very high).

8.3 Benchmarking Company Performance Using the SOM Model

In the following, an example of actual benchmarking will be illustrated. More examples of the type of benchmarking that can be performed can be found in Paper 2. Some other results are also briefly presented in Paper 3. Where interesting observations are made, explanations for these will be sought with help of the information in the companies' annual reports. This process will help to determine the overall performance of the self-organizing map. The benchmarking has been performed using trajectories (see Section 6.3.5)

8.3.1. Benchmarking the Top 5

According to the Finnish Forest Industries Federation (2004), the largest global pulp and paper companies according to net sales in 2003 were International Paper, Georgia-Pacific, Weyerhaeuser, Kimberly-Clark, and Stora Enso. As an example of performance benchmarking, these five companies will be benchmarked against each other. These results are an updated version of the results presented in Paper 3. The results are displayed in Figure 8.4 and Figure 8.5.

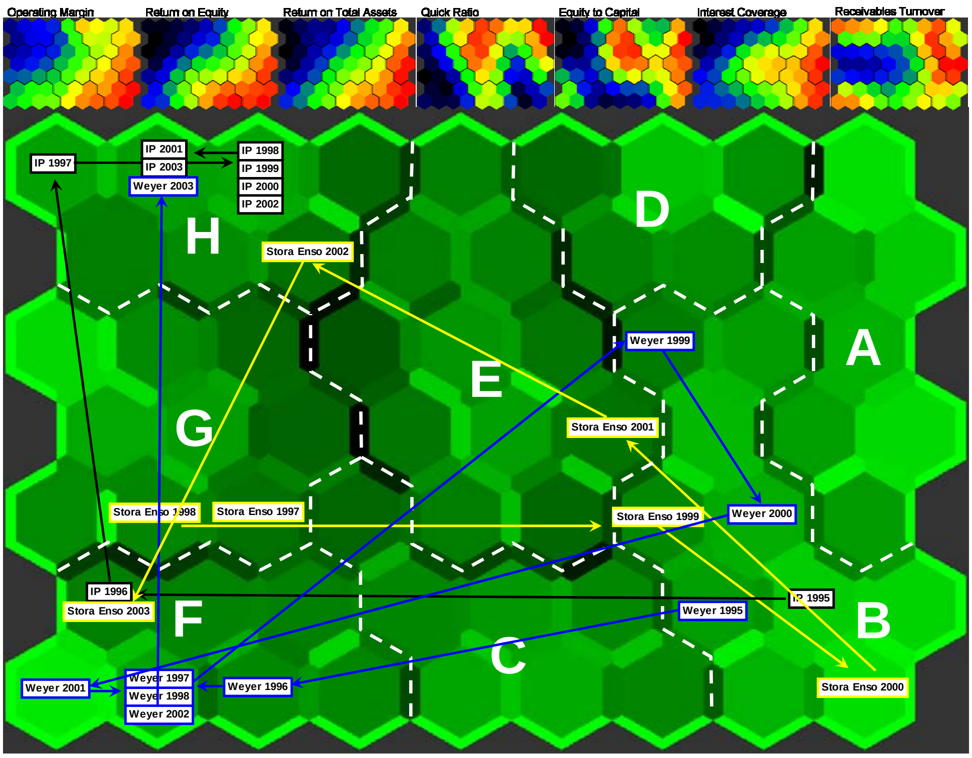


Figure 8.4. International Paper, Weyerhaeuser, and Stora Enso from 1995-2003.

An interesting note is that International Paper (Figure 8.4), the largest pulp and paper manufacturer in the world for several years, consistently shows weak performance, particularly according to profitability ratios. In 1995, when the industry in general was doing very well, in particular due to high market pulp and paper prices, IP is located in Group B, and its performance was strong. In 1996, the company falls to Group F, primarily due to falling market prices and excess supply in the industry. The same trends can be seen in many of the Top 5 companies. In the following years, reduced profitability forces the company into Group H, indicating very low profitability and solvency. However, efficiency and liquidity have improved somewhat from 1996. IP's performance is characterized by several years of negative earnings, in particular due to extraordinary costs in the form of restructuring costs. In 2002 and 2003, IP suffered from the difficult economic situation in the US, Asia, and Europe, something that also shows in the other Top 5 companies' results.

Weyerhaeuser (Figure 8.4), third largest in the world, again shows the effects of the favorable market situation in 1995, and also, the drop in prices during 1996. In 1999, record net sales and good prices across their entire range of products attributed to the good financial results. Weyerhaeuser also acquired MacMillan Bloedel during the last quarter of 1999. Results continued to improve in 2000, again with record net sales. In 2001, decreased prices, closures of mills, acquisition costs for MacMillan Bloedel, and the generally poor industry conditions heavily affected performance in 2001, dropping the company into Group F. In 2003, profitability increased somewhat, but solvency dropped, pushing the company into the poorest group.

Stora Enso (Figure 8.4), the fifth largest in the world, emerged in 1998 when Stora (Sweden) merged with Enso (Finland). Consolidated statements were used to illustrate the company in 1997. In 1998, the significant costs of the merger placed the company in a very unprofitable group, Group G (low profitability and solvency). In 1999, Stora Enso's performance improves considerably, moving the company into Group B. In 2001, Stora Enso's profitability fell somewhat, dropping the company into Group E. Profitability in 2001 is average, but solvency is very high. In 2002, the results were heavily influenced by a very large write-down in the asset value of Consolidated Papers, acquired in 2000. The one-time charge, coupled with difficult conditions in the industry, pushed Stora Enso into the poorest group. 2003 was a particularly difficult year for European pulp and paper companies, primarily due to the weak dollar. The consequence of the weak dollar is that European products become very expensive, as paper products are generally priced in USD. The results are evident, in particular, in Finnish pulp and paper companies' annual results. However, Finnish companies' efficiency is still much better than that of most of their American competitors'.

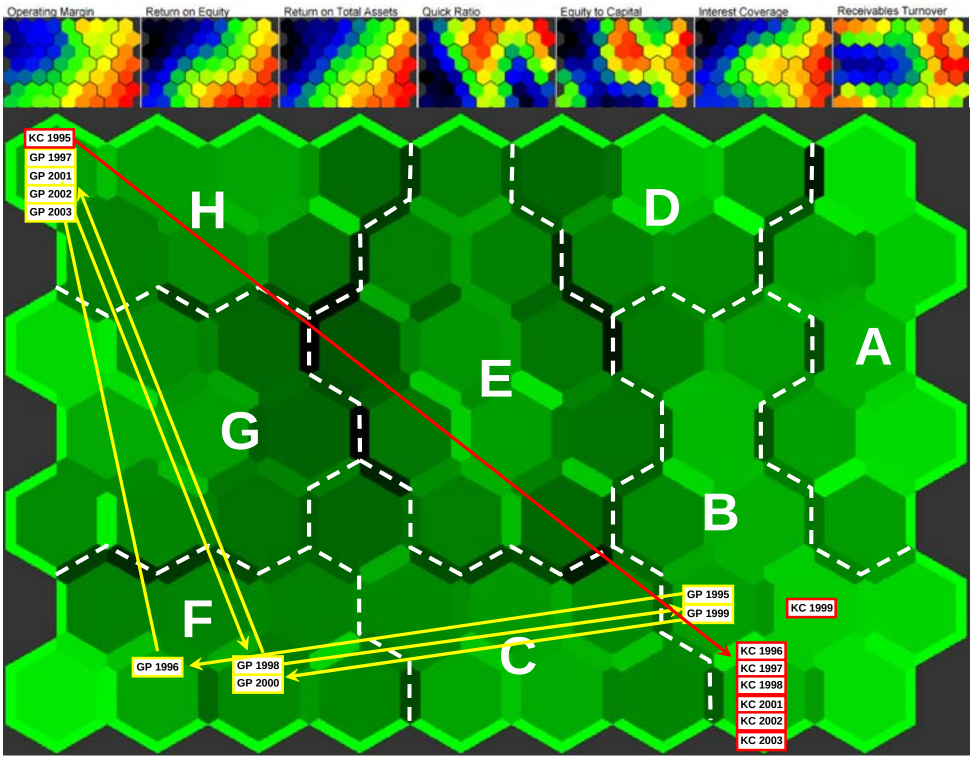


Figure 8.5. Georgia-Pacific and Kimberly-Clark from 1995-2003.

Georgia-Pacific (Figure 8.5), the second largest in the world, rose into the Top 5 through its acquisition of Fort James. With the exception of during 1995 and 1999, GP's performance is not very convincing. Profitability is average to weak, and solvency is very weak. The company can be found in Group B during two years (1995 and 1999), but for other years it can be found in Groups F or H. During 1995 the good results were primarily due to the generally favorable market situation (average market pulp prices were up 70% from 1994) as well as the sale of assets. However, prices started declining by the end of 1995, and the results in 1996 were considerably poorer for the same reasons as for IP. In 1999, results were again up due to the sale of timberland, a pretax income of 355 million USD. Also included was a recent acquisition. In December 2000, Georgia-Pacific acquired Fort James. The acquisition was significant, and therefore already had an effect on the financial results of the same year, in particular due to the increased debt. The effect was much higher in 2001, causing negative results and moving the company into Group H. The effects of the high level of debt can be seen in that interest expenses doubled in 2001 compared to 2000. Since 2001, difficult conditions in the industry have kept GP in the poorest group.

Kimberly-Clark (Figure 8.5), fourth largest in the world, is a very strong performer and industry benchmark. During 1995, Kimberly-Clark's performance was burdened by one-time charges relating to the acquisition of its formerly main competitor, Scott Paper. Kimberly-Clark has become a favorite management textbook example of effective management and leadership since the early nineties. The dramatic changes resulting in the acquisition of Scott Paper are often attributed to the leadership of Kimberly-Clark's CEO, Darwin Smith (Collins 2001). Kimberly-Clark's performance remains very strong and stable throughout the entire experiment.

In conclusion, one can note that Kimberly-Clark is clearly the strongest performer in the Top 5. The company displays stable, strong performance for the duration of the experiment, with the exception of 1995. Stora Enso is clearly more efficient than its larger US competitors, although Stora suffered during the difficult years in the industry. On the other hand, in 2003 Stora Enso's performance is considerably better than the US companies'. The US companies clearly show that there is a considerable benchmarking gap between the efficient Finnish companies and large US companies. The benchmarking has also shown that 1995 and 2000 were generally good years for the Top 5 companies, while 2001-2003 have been poor.

8.4 Combining Company and Macro Level Analysis

In Paper 4, the concept of multilevel environment analysis using the SOM is proposed and illustrated. In this section, a brief summary of the results is presented. In the paper, two SOM models (pulp and paper industry-level and firm-level maps) are compared in order to identify simultaneous changes on both levels.

The paper begins with an introduction to multilevel analysis and benchmarking, and discusses the importance of performing multilevel comparisons.

Two SOM models are created for the experiment. The first model deals with industry level indicators in eight countries: Austria, Canada, Finland, France, Germany, Japan, Sweden, and the United States. The ten variables included are total input price, labor price, raw material price (cost indicators, all in FIM), total production quantity, total productivity, productivity of raw material, productivity of energy, productivity of work, and productivity of capital (productivity indicators). The second model, a prior model of the model presented in this thesis, is a firm level financial comparison similar to this model. The period of comparison was 1995-2000. The study had two purposes: to a) analyze whether the development of productivity and costs affect companies' financial ratios and

b) to illustrate how the industry-level factors affect the largest pulp and paper companies. The analysis was performed using both country-level averages calculated based on the companies from each country, as well as the individual companies themselves.

There were a number of conclusions based on the study. Firstly, Finnish and Swedish companies were identified as above average performers. Generally, these companies displayed higher profitability than their competitors. On the industry level, Finnish and Swedish costs were lower than in other countries, while productivity was higher. This partly explained these countries' better than average performance. However, market pulp prices (not included in the study) were found to have strongly affected the performance of these countries.

In the US, on the other hand, costs were very high during the period, particularly during 1997-2000. This helps to explain the poor performance of US companies during this period. The same applies to Canadian companies, which are heavily affected by the sharp rise in costs during 1997. However, rising productivity leads to improvement in the Canadian ratios in 2000.

The reason for the poor performance of Japanese companies is evident on the industry level. Costs are extremely high compared to other countries, while productivity remains low. The effect of the Asian financial crisis is clear on both levels.

The same analysis is performed in an attempt to assess the reasons for individual companies' financial performance. As on the country level, the effect of some of the industry level indicators shows up in the performance of the individual companies.

In conclusion, the study shows the importance of multilevel performance analysis. In the study, the connection between general industry level factors and individual company performance are shown. These indicators are important to take into consideration when the performance of an individual company is assessed. However, the study also noted that not all changes were possible to explain using the maps, indicating a need for further research in the area. Finally, macro level analysis was proposed as an extension to the analysis.

8.5 Combining Quantitative and Qualitative Analysis

In Paper 5, the addition of qualitative analysis is discussed. The paper describes the combination of two methods for the analysis of financial performance: the SOM for analyzing quantitative financial ratios and a prototype-matching text mining approach for analyzing the textual parts of annual reports. The motivation for the research was the assumption that quantitative financial ratios can be seen as historical facts (what has happened), while the text contains explanations for the events (why it happened) as well as non-obvious indications for the future. In addition, the study illustrates the use of the SOM for analyzing the quantitative parts of quarterly reports.

For the study, quarterly reports for companies in the international telecommunications industry were collected, and a SOM model was trained as in this dissertation. The financial ratios used were the same, except for that the current ratio was used instead of quick ratio to measure liquidity. In addition, the data had to be scaled in order to be comparable to annual level figures, as the balance sheet in quarterly reports in for the past twelve months, whereas the income statement is for the past three months. The studied period consisted of 2000-2001. A financial benchmarking of the top three telecommunications companies was performed, which identified Nokia as the strongest performer.

In addition, a prototype-matching text mining method (Visa et al. 2002; Kloptchenko 2003) was used to mine the textual parts of the quarterly reports. The prototype-matching algorithm is a text mining method based on word and sentence level encoding and clustering. The idea is that it clusters documents based on similarities in the text, and returns the closest matches to a specified document (a prototype). A technical description can be found in Visa et al. (2002) and Kloptchenko (2003). The textual clustering was performed for Nokia, Ericsson, and Motorola, for the period 2000-2001.

The results of the textual clustering were compared to the results of the financial clustering, in order to see if there is a correlation between the content of the report and the reported financial performance.

Generally speaking, a tendency for the results to indicate future performance was noticed. It was noticed that the closest matches to a prototype indicate a general level of performance for the following quarter. In Table 8.7 an example of the results is presented. The heading of each column indicates the prototype report and the letter indicates the level of performance in that quarter. The clusters of the financial benchmarking are in order of decreasing attractiveness, i.e. A1 – A2 – B – C1 – C2 – D. For example, the first column indicates that the prototype report for Ericsson quarter 1, 2000 is located in cluster B in the financial benchmarking. The closest matches to the report are Nokia quarter 1, 2000 (cluster A1) and Nokia, quarter 3, 2000. In the following quarter, Ericsson's performance increases to cluster A1. The following closest match is from cluster B, after which performance drops back into cluster B. After this, poorer reports start appearing, as Ericsson's performance continues to worsen.

Ericsson00Q1 B	Ericsson00Q2 A_1	Ericsson00Q3 B	Ericsson00Q4 C_1	Ericsson01Q1 C_1	Ericsson01Q2 D
Nokia00Q1 A_1	Ericsson00Q3 B	Ericsson00Q4 C_1	Ericsson00Q3 B	Ericsson01Q2 D	Nokia01Q3 C_2
Nokia00Q3 A_1	Nokia00Q2 A_1	Motorola01Q3 D	Motorola01Q2 C_2	Ericsson01Q3 D	Ericsson01Q1 C_1
Motorola01Q3 D	Ericsson00Q1 B	Ericsson00Q2 A_1	Motorola01Q3 D	Nokia01Q3 C_1	Ericsson01Q3 D
Motorola01Q2 C_2	Ericsson00Q4 C_1	Ericsson00Q1 B	Nokia00Q1 A_1	Motorola01Q3 D	Nokia01Q1 A_1

Table 8.7. Example of combination of quantitative and qualitative clustering.

Generally speaking, the trend for Ericsson was quite nicely captured, but it was not as nicely captured for Motorola or Nokia. Because of the small sample used, as well as some problems associated with early versions of the method (Kloptchenko 2003), it is difficult to judge the degree of confidence in the results. However, to this thesis, the technical development of the prototype matching algorithm is not in itself relevant, as it is the concept of combining data and text mining methods for additional insight that is important. Indeed, any suitable text mining method could be combined with the SOM in such an application.

The conclusions of the research indicate the potential value of combining quantitative and qualitative analysis in financial comparisons. It is important to note that we are concerned with identifying non-obvious indications, not making predictions. The difference lies in that the textual parts of annual reports already contain some information about future prospects, as these are known to managers. For example, in most industries managers already know the following quarter's order stock, an essential figure relating to coming financial performance. The text on future prospects will reflect this information in one way or another. Using a combination of data and text mining methods offers the possibility of automating the search process to a certain degree.

Similar research, but using a linguistic method instead of text mining, was explored in Paper 6. The linguistic method used was the collocational network, which is a method that looks for words in a text that appear within a specified distance (window) of each other a significant amount of times in a text. The idea is that some information about the content of a text can be revealed by simply looking for words that consistently appear together a text. Often, the collocates bear some implied meaning, such as positive or negative undertones (for example, increased – sales, efficiency – program, decreased – earnings, etc.). In paper 6, collocational networks are combined with SOM analysis to study the same telecom database as in paper 5. It was possible to find a trend in the changes in the structure of the collocational networks, before an actual change in the financial performance was seen on the SOM map. As in paper 5, combined quantitative/qualitative analysis was found to yield more potential value than either method alone. However, the dataset was limited, and collocational networks are difficult to use in larger scale problems. Human interpretation is necessary, and collocational networks are thus difficult to automate.

8.6 Summary

In this chapter, the training and analysis of the map have been discussed. First, the characteristics of the data and the required preprocessing methods were determined. Then, the training of the map was discussed based on recommendations in the literature, and the results were displayed. The map was analyzed, and a benchmarking of the Top 5 companies was performed. The connection between industry level factors and financial performance was shown, and the importance of multilevel analysis was demonstrated. Finally, the application of qualitative analysis was introduced as a potential improvement to the model. In the next chapter, an evaluation of the results will be performed.

9 EVALUATION OF THE SOM MODEL

9.1 Background

The objective of the expert evaluation was to perform a face validation of the proposed SOM model. The face validation was to be performed by subject matter experts (SMEs) in the area of competitor analysis, i.e. business intelligence managers, financial managers, and corporate development managers. The purpose of the survey was to validate the benchmarking model.

9.2 Research design

For the expert evaluation of the model, demos of the SOM models were arranged at the premises of the participating companies. The demo consisted of three parts: a 15 minute presentation of the basics of the SOM, a 35 minute presentation of the financial benchmarking model, and a 35 minute presentation of the macro environment model. The macro environment model is beyond the scope of this thesis, more information can be found in Länsiluoto et al. (2002).

After the presentation, the participants were given a questionnaire to complete. The questionnaire was again based upon the Doll and Torkzadeh model, modified to fulfil the requirements of this experiment. Specifically, the factors included in the model were further divided into several sub factors, in order to better be able to evaluate the different aspects of the quality of the model. The factors are viewed from an information quality perspective, i.e. how useful the information provided by the model is to potential decision makers (Alter 2002, p.163).

Most of the questions were based on a 5 point Likert scale (1 = strongly disagree, 2 = somewhat disagree, 3 = neutral, 4 = somewhat agree, 5 = strongly agree). There were also a number of other attitude scales, such as very important – very unimportant, very satisfied – very dissatisfied, etc., also on 5 point scales. Finally, there were a number of open questions. The questionnaire was administered in English, as most of the companies were multinational in their operations.

9.3 Sample and administration

In the state of the art survey (see Section 3.2), respondents were queried concerning their interest in taking part in a SOM demo and evaluation. 15 of the respondents expressed preliminary interest in the evaluation. Of these, 13 were visited. The two remaining companies choose to not participate due to time constraints. The participants consisted of 11 members of industry, including two major pulp and paper companies (totalling 14 responses, i.e. a significant proportion), and two from banking and investment. The number of participants from each company varied between one and seven. The total number of questionnaires distributed was 39, and the total number of responses was 36. The response rate was thus 92.31%. The sample size conforms to Roscoe's (1975) rule of thumb recommendation for minimum sample size (30) for statistical analysis (Hill 1998). The experiment was conducted between January and March 2004.

9.4 Results

Before analyzing the data, it is important to note a number of assumptions and limitations.

First of all, when working with attitude scales it is important to note that all respondents might not perceive the "distances" between two answers as the same. For example, for one user the difference between "somewhat agree" and "strongly agree" might be larger than for another. As it has been argued that users can rarely discriminate between seven possible answers on an attitude scale (Viswanathan et al. 2004), 5-point Likert scales were chosen. Using a 5-point scale should result in more uniform answers than, for example, using a 7-point scale. Viswanathan et al. (2004) argue for a 3-point scale, but this would probably have provided too little differentiation in the answers. Under the assumption that these distances are uniform, researchers commonly treat attitude scales as interval scales (Tull and Hawkins 1987, p.216), instead of ordinal scales. Ordinal scales do not allow the use of parametric statistics. Therefore, for the purpose of statistics, it has been assumed that the distances between categories are uniform and can be treated as interval scales.

Secondly, for the purposes of testing the hypotheses, the average of each factor of information has been calculated. Thus, it is assumed that each of the sub factors (for example, relevance, informativeness, importance, helpfulness, and sufficiency, for content) combined are a valid measure of each factor, and that they are all equal in importance. Under this assumption, individual items can be summed to indicate an overall score for an individual (Tull and Hawkins 1987, p.297).

Thirdly, bias becomes an important factor in face validation settings. Two types of bias in particular are important to note: a bias (positive or negative) towards the interviewer, and a bias towards answering what the interviewer wants to hear. In the first case, the respondent's answers may be affected by their perception of the interviewer based on appearance, race, gender, etc., and in the second case, the respondent modifies his or her answers based on what he or she perceives that the interviewer wants to hear. Although the questionnaire was self-administered, the presence of the researchers may have had an affect on the respondents' answers. Unfortunately, there was no other option for administering this face validation. The 'the error of central tendency' might also be a factor, i.e. respondents tend to not use the extreme ends of scales.

Finally, all of the factors are not entirely possible to directly evaluate in a face validation setting. In a face validation approach, the evaluation of the artefact is based on the respondent's perception of the features of the artefact. This is especially true in the case of ease of use and timeliness in this example, as the managers are not able to directly evaluate these features by manipulating the model. Thus, the results may be influenced by the presentation of the model. However, a strong effort was made to provide each company with a presentation that was as similar as possible. This is important to bear in mind during the analysis of the data.

9.4.1. Demographics

Background information was optional, but 91.67% of the respondents provided this information. 81.25% of those who provided background information had a master's of science degree, while 6.06% had a higher academic degree. Therefore, the education level of the respondents was very high. Of those who provided information about their current position, 81.25% were directly involved with business intelligence, strategic development, analysis, or corporate finance.

The average number of years in the current company was 8.16 years (standard dev. 8.90), with an average of 3.75 years in the current position (standard dev. 4.10), and 5.8 years in a similar position in the manager's career (standard dev. 6.60).

Compared to the respondents in the state of the art survey, the managers participating in the evaluation had higher level of education (87.85% with a master's degree or higher, versus 78.57% with a master's or higher in the first phase). The targeted population was, as expected, slightly more effectively reached in the evaluation.

Familiarity with IT tools	N	Mean	SD	Med	Very Experienced – Neutral – Very Inexperienced 5	4	3	2	1
Word processing	35	3.80	0.53	4	5.71%	**68.57%**	25.71%	0.00%	0.00%
Spreadsheets	35	3.89	0.58	4	11.43%	**65.71%**	22.86%	0.00%	0.00%
E-mail	34	4.00	0.49	4	11.76%	**76.47%**	11.76%	0.00%	0.00%
Calendars	35	3.57	0.61	4	2.86%	**54.29%**	40.00%	2.86%	0.00%
Databases	35	3.09	1.01	3	8.57%	22.86%	**42.86%**	20.00%	5.71%
Internet	35	3.97	0.62	4	17.14%	**62.86%**	20.00%	0.00%	0.00%
Decision support systems	32	2.31	1.03	2	3.13%	6.25%	**34.38%**	31.25%	25.00%

Table 9.1. Familiarity with IT tools.

The degree of IT familiarity of the managers was again high (Table 9.1). The majority used word processing (77.78%), spreadsheets (69.44%), e-mail (100.00%), calendars (80.56%), and the Internet (100%) daily. The only significant difference in the use of IT tools was the much higher reported use of electronic calendars during the state of the art survey. Apart from this, the IT demographics of both groups are quite similar. Interestingly, the second group reported a slightly more frequent use of decision support systems, but at the same time, a slightly lower experience in these. The difference, however, was not significant according to an independent samples t-test.

We can again conclude that the managers are experienced users of basic IT applications, and are therefore familiar with computers. However, few are experienced users of more than basic applications. The demographics of the managers participating in the evaluation are thus very similar to the demographics of the managers in the state of the art survey.

9.4.2. Current methods

Importance of factors of information	N	Mean	SD	Med	Very important – Neutral - Very unimportant 5	4	3	2	1
Content	36	4.47	0.56	4.5	**50.00%**	47.22%	2.78%	0.00%	0.00%
Accuracy	36	4.33	0.63	4	41.67%	**50.00%**	8.33%	0.00%	0.00%
Format	36	3.64	0.80	4	11.11%	**50.00%**	30.56%	8.33%	0.00%
Ease of use	36	4.08	0.69	4	25.00%	**61.11%**	11.11%	2.78%	0.00%
Timeliness	36	4.08	0.65	4	22.22%	**66.67%**	8.33%	2.78%	0.00%

Table 9.2. Importance of factors of information.

An independent samples t-test shows that the differences between the importance ratings of factors of information achieved in the evaluation phase (Table 9.2) do not differ significantly from those reported in the state of the art survey (paper 1). The most important factors are content and accuracy, while format is again the least important. Therefore, it can be concluded that the managers participating in the second phase prioritize the same factors as the mangers in the first phase. Again, the Doll and Torkzadeh model appears to provide a good framework for this study.

Satisfaction with current methods	N	Mean	SD	Med	Very satisfied – Neutral – Very dissatisfied 5	4	3	2	1
Content	36	2.89	1.01	3	2.78%	**30.56%**	25.00%	36.11%	5.56%
Accuracy	36	2.94	0.98	3	2.78%	**30.56%**	**30.56%**	**30.56%**	5.56%
Format	36	2.97	0.84	3	0.00%	27.78%	**47.22%**	19.44%	5.56%
Ease of use	36	2.69	0.89	3	0.00%	16.67%	**47.22%**	25.00%	11.11%
Timeliness	36	3.11	0.95	3	5.56%	27.78%	**44.44%**	16.67%	5.56%

Table 9.3. Participants' satisfaction with current methods for financial benchmarking.

The results in Table 9.3 show that the participants were significantly (independent samples t-test) less satisfied with the content ($t(72) = 2.21$, $p = 0.03$), accuracy ($t(72) = 1.72$, $p = 0.09$) and ease of use ($t(72) = 2.40$, $p = 0.019$) factors of current methods than in the state of the art survey (Table 3.2). This could be due to two reasons. Firstly, the participants in the state of the art survey who were less pleased with their current methods might have been more willing to take part in the evaluation. Secondly, participants in the evaluation might have been less pleased with current methods after having viewed the alternative method. An independent samples t-test shows that the average of the identifiable companies in the state of the art survey (N = 9), who also took part in the evaluation, only differs significantly from the average satisfaction (N = 36) of the evaluation in the content factor ($t(26.30) = 3.29$, $p = 0.003$). Although the results are not entirely comparable due to the large difference in the number of samples on each side (do not conform to Roscoe's (1975) rules of thumb, as the sub-samples are smaller than 30), it does lend some support for the first option, and thus viewing, the model has not strongly affected the participants' previous perceptions of their own methods. Instead, it appears that they were indeed less satisfied than average from the beginning.

	N	Mean	SD	Med	Very complex – Neither – Very uncomplex 5	4	3	2	1
Complexity of the competitive environment	35	3.74	1.07	4	25.71%	**42.86%**	11.43%	20.00%	0.00%
	N	Mean	SD	Med	Constantly frustrated – Neither - Never frustrated 5	4	3	2	1
Frustration with daily information	35	2.97	1.01	3	0.00%	**42.86%**	17.14%	34.29%	5.71%
	N	Mean	SD	Med	Very turbulent – Neither - Not turbulent 5	4	3	2	1
Turbulence of the competitive environment	35	3.60	1.09	4	20.00%	**45.71%**	8.57%	25.71%	0.00%

Table 9.4. Characteristics of the competitive environment.

Table 9.4 shows how the managers perceive the competitive environment. An independent samples t-test shows that the results do not differ significantly from those achieved in the state of the art survey. Managers do perceive a high complexity and turbulence in the competitive environment, but they are not particularly frustrated by the amount of information they are facing daily. However, it can again be concluded (as in Section 3.2) that roughly half of the managers often face information overload. It can also be concluded that the complexity of the competitive environment is high, and that complexity reducing tools could, therefore, be valuable for business intelligence managers. Potentially, the SOM could, thanks to its dimensional reduction and visualization capabilities, provide such a tool. Support for this suggestion can be found in Table 9.14.

9.4.3. Evaluation of the model

Content

Content of the model	N	Mean	SD	Med	Strongly agree – Neutral – Strongly disagree 5	4	3	2	1
Relevant	36	4.03	0.65	4	19.44%	**66.67%**	11.11%	2.78%	0.00%
Informative	36	3.86	0.72	4	16.67%	**55.56%**	25.00%	2.78%	0.00%
Important	36	4.00	0.68	4	22.22%	**55.56%**	22.22%	0.00%	0.00%
Helpful	36	4.03	0.81	4	27.78%	**52.78%**	13.89%	5.56%	0.00%
Sufficient	36	3.17	0.85	3	0.00%	**41.67%**	36.11%	19.44%	2.78%

Table 9.5. Content of the model.

Overall, the content of the model has received a very high rating from managers (Table 9.5). However, the sufficiency of the model has received considerably lower ratings, and was the only characteristic to receive any strongly disagree answers (2.78%). Obviously, the managers feel that the model is lacking something. This was not unexpected, as this is a very general benchmarking model. Most companies have their own combination of measures that they routinely use, so any general model will lack something. One of the things that managers may be lacking is the addition of qualitative analysis, something that is explored in Paper 5. This conclusion is also supported by some of the answers to the open questions. However, the number of managers who agree is still twice as high as those who disagree.

We can conclude that the managers are quite satisfied with the content of the model.

Accuracy

Accuracy of the model	N	Mean	SD	Med	Strongly agree – Neutral – Strongly disagree				
					5	4	3	2	1
Reliable	36	3.56	0.77	4	5.56%	**55.56%**	27.78%	11.11%	0.00%
Precise	36	3.11	0.89	3	0.00%	**41.67%**	30.56%	25.00%	2.78%
Valid	36	3.81	0.62	4	8.33%	**66.67%**	22.22%	2.78%	0.00%
Complete	36	2.92	0.91	3	0.00%	30.56%	**36.11%**	27.78%	5.56%
Overall, accurate	36	3.44	0.73	4	0.00%	**55.56%**	36.11%	5.56%	2.78%

Table 9.6. Accuracy of the model.

Accuracy has also achieved high ratings (Table 9.6), with the exception of two characteristics, precision and completeness. The number of managers who agree that the model is precise is higher than the number who disagree, but it is obvious that there is room for improvement in the precision of the model. The somewhat lower ratings in precision seem to indicate a lack of drill-down facilities, something that is already available in some commercial SOM applications (although there are fairly few of these). Although the managers appreciate the quick overview afforded by the SOM, they would also apparently like to obtain crisp numbers for each company, something that should be considered when creating a testable prototype of the model.

The score in completeness is somewhat difficult to judge. Complete refers to "*having all necessary parts, elements, or steps*"[13], i.e. in terms of accuracy, is there something missing that would significantly affect the result? Sufficiency

[13] Merriam-Webster Online Dictionary, http://www.m-w.com

refers to “*enough to meet the needs of a situation or a proposed end*”, i.e. in terms of content, does it fill the requirements of financial benchmarking? However, this may have caused confusion among the managers, especially as the questionnaire was administered in English and most of the managers were Finnish. In either case, the score is quite low compared to the other aspects of accuracy.

When asked if the models appear to correlate with reality, 93.94% replied yes (27.27% of these replied very similar). Overall, the managers are also satisfied with the accuracy of the model.

Format

Format of the model	N	Mean	SD	Med	Strongly agree – Neutral – Strongly disagree 5	4	3	2	1
Satisfactory colors	36	3.92	0.81	4	19.44%	**61.11%**	11.11%	8.33%	0.00%
Satisfactory shapes	36	3.83	0.65	4	11.11%	**63.89%**	22.22%	2.78%	0.00%
Clear visual representation	36	3.56	1.00	4	16.67%	**41.67%**	22.22%	19.44%	0.00%
Readable maps	36	3.47	0.77	4	5.56%	**47.22%**	36.11%	11.11%	0.00%
Overall, the format is satisfactory	35	3.77	0.73	4	14.29%	**51.43%**	31.43%	2.86%	0.00%

Table 9.7. Format of the model.

The format (Table 9.7) was considered satisfactory according to all measures by the majority of the managers. In addition, no managers strongly disagreed with any of the statements, but many strongly agreed (colors, 19.44%; shapes, 11.11%; clear representation, 16.67%; and overall satisfaction, 14.29%).

We can therefore conclude that the managers were very satisfied with the format of the models. This is an important conclusion, as the format of the information presented in the SOM model is arguably its main contribution. This is valuable support for the research presented in this dissertation.

Ease of Use

Transparency	N	Mean	SD	Med	Very transparent – Neutral – Very non-transparent 5	4	3	2	1
Transparency of the model	36	4.00	0.76	4	19.44%	**69.44%**	2.78%	8.33%	0.00%

Table 9.8. Transparency of the model.

Using the SOM, it is easy to perceive and analyze:	N	Mean	SD	Med	Strongly agree – Neutral – Strongly disagree 5	4	3	2	1
comparable data	36	4.08	0.60	4	22.22%	**63.89%**	13.89%	0.00%	0.00%
data trends	36	4.03	0.70	4	19.44%	**69.44%**	5.56%	5.56%	0.00%
correlations between variables	36	3.75	0.87	4	16.67%	**52.78%**	19.44%	11.11%	0.00%
data clusters	36	4.31	0.58	4	36.11%	**58.33%**	5.56%	0.00%	0.00%
differences between data	36	3.39	0.80	3	5.56%	**41.67%**	38.89%	13.89%	0.00%
data values	36	3.08	0.77	3	0.00%	33.33%	**41.67%**	25.00%	0.00%

Table 9.9. Ease of use of the model.

The SOM can be conveniently used by:	N	Mean	SD	Med	Strongly agree – Neutral – Strongly disagree 5	4	3	2	1
a) an expert user	35	4.43	0.70	5	**51.43%**	42.86%	2.86%	2.86%	0.00%
b) an end / business user	36	2.86	0.87	3	0.00%	27.78%	33.33%	**36.11%**	2.78%

Table 9.10. Ease of use of the SOM in general.

The ease of use of the model has been judged using three measures, transparency of the model (Table 9.8), perceived ease of use for different tasks (Table 9.9), and perceived technical ease of use (Table 9.10). As can be seen from the tables, a very strong majority of the managers agree that the model is transparent (easy to understand). The SOM is also perceived to easily support a number of different analysis tasks. However, differences between data and in particular data values received considerably lower scores, although not poor. This correlates with the precision rating in the accuracy factor (Table 9.6), which was also somewhat lower than overall accuracy. It does appear that stronger drill-down capabilities would have been valued by the managers.

The managers did perceive that the SOM can be conveniently used by experts; however, most managers also feel that the use of the SOM requires expert experience. A possible conclusion, supported by many of the answers to the open questions, is that the managers feel that the SOM would be most suitable for fairly infrequent use, such as during the strategy process. A particular strength noted by the managers was the ability to obtain a quick overview of the competitive environment (Table 9.14). They also commented that reading the map requires a certain amount of experience, which makes it preferable that the SOM is used continuously, not as a one-time analysis.

Therefore, it can be concluded that from an information perspective, assuming that the SOM expert is proficient in training and presenting the model, the model is easy to use. However, a SOM expert is required for the presentation of the results.

Timeliness

Timeliness of the model	N	Mean	SD	Med	Strongly agree – Neutral – Strongly disagree 5	4	3	2	1
The model is timely	36	3.14	0.87	3	0.00%	**41.67%**	33.33%	22.22%	2.78%

Table 9.11. Timeliness of the model.

The timeliness of the model (Table 9.11) received an average rating by the managers. A possible reason for this is that the model is dependent upon the same time constraints as any other benchmarking method or model, i.e. access to new data. Therefore, the timeliness factor is perhaps not as relevant to this evaluation as the other factors are.

Overall satisfaction

Overall satisfaction with the SOM	N	Mean	SD	Med	Very satisfied – Neither – Very dissatisfied 5	4	3	2	1
Overall satisfaction	35	4.03	0.71	4	25.71%	**51.43%**	22.86%	0.00%	0.00%

Table 9.12. Overall satisfaction with the model.

The overall satisfaction with the model among the managers was very high. 25.71% reported that they were very satisfied with the SOM model, and no managers were dissatisfied. Thus, the model received very strong support from the managers.

9.4.4. Information Usage

Information usefulness	N	M	SD	Med	Strongly agree – Neutral – Strongly disagree 5	4	3	2	1
Information in the form of the models presented could improve the quality of a strategic decision (e.g. investment planning)	36	3.83	0.91	4	16.67%	**63.89%**	8.33%	8.33%	2.78%
Information in the form of the models presented could improve confidence in a strategic decision (e.g. investment planning)	36	3.53	0.94	4	11.11%	**47.22%**	27.78%	11.11%	2.78%
Information in the form of the models presented could affect or stimulate discussion during the strategic process.	36	4.33	0.72	4	44.44%	**47.22%**	5.56%	2.78%	0.00%
I would use the models if they were presented to me in the strategy process.	36	3.94	0.89	4	30.56%	**38.89%**	25.00%	5.56%	0.00%

Table 9.13. The SOM in strategic planning.

The managers were also asked how they would perceive the usefulness of the SOM model in strategic decision making settings. The results (Table 9.13) indicate that this is an area in which the managers perceive high value in the SOM, as was indicated earlier. A majority of the managers agreed with the statements concerning the use of the SOM in strategic decision making. The answers to the questions in Table 9.13 suggest that the managers perceive value in using the SOM to stimulate discussion and obtain overviews during the strategy process, i.e. infrequent but consistent use. As was noted in the section on ease of use, the managers perceive more value in this type of use than in daily use.

	N	Mean	SD	Median	Mode			
	N				Mode	Yes	Not sure	No
Could our models be helpful in the handling of the competitive environment?	36				1	**75.00%**	19.44%	5.56%
	N	Mean	SD	Median	Mode	Well	Neither	Poorly
Did the SOM show investments, turnarounds, problems and new moves well?	35	3.74	0.78	4	4	**71.43%**	20.00%	8.57%
	N	Mean	SD	Median	Mode	Helpful	Neither	Unhelpful
Did our models help you to obtain a quick overview of the competitive environment?	36	4.19	0.67	4	4	**91.67%**	5.56%	2.78%
	N				Mode	Yes	Not sure	No
Did our models provide any new information about the competitive environment?	36				1	**52.78%**	16.67%	30.56%
	N	Mean	SD	Median	Mode	Some additional	Neutral	Few additional
Does the SOM provide additional benefits over currently used analysis methods?	35	4.06	0.76	4	4	**85.71%**	8.57%	5.71%

Table 9.14. Questions concerning dealing with complexity using the SOM model.

Table 9.14 provides support for the suggestion of using the SOM for dealing with complexity in the competitive environment. The managers obviously perceive the SOM to be a good tool for gaining a quick overview, as is indicated by results in the table.

					Absolutely – Undecided - Absolutely not				
Use of the SOM	N	Mean	SD	Med	5	4	3	2	1
Would you use the SOM as a complement to other tools in analysis cases such as those demonstrated?	35	3.94	0.68	4	17.14%	**62.86%**	17.14%	2.86%	0.00%
Could the SOM replace one or more of the tools currently used in analysis cases such as those demonstrated?	35	3.20	0.87	3	0.00%	**45.71%**	31.43%	20.00%	2.86%
Would you recommend the SOM to your colleagues?	35	3.80	0.76	4	20.00%	**40.00%**	**40.00%**	0.00%	0.00%

Table 9.15. Conclusions concerning the use of the SOM.

The concluding questions (Table 9.15) again provide support for the use of the SOM in financial benchmarking. Over 45% of the managers thought that the SOM could possibly replace one or more of their currently used tools, and the managers also strongly indicated that they would recommend the SOM to their colleagues. However, there was also a group of over 20% of the managers that replied that the SOM could probably not replace current methods in the company. As was indicated by a number of the answers to the open questions, the lack of qualitative analysis methods might in part explain this hesitation. In general, however, the support for the method is strong.

9.5 Satisfaction with the model

In order to validate the model, a number of hypotheses have been posed. The hypotheses are posed to validate the model according to the five factors of information proposed by Doll and Torkzadeh. As was mentioned in Section 9.4, per respondent-averages of the sub factors were calculated for each factor, in order to obtain an overall score in each factor for each respondent. This can be done under the assumptions presented in section 9.4 (Tull and Hawkins 1987, p.297).

Thus for each factor, the following hypothesis, where <<factor>> refers to each of the factors in the Doll and Torkzadeh model (content, accuracy, format, ease of use, and timeliness), was proposed:

H_0: *Respondents did not rate the <<factor>> of the SOM model higher than their own methods*

H_1: *Respondents rated the <<factor>> of the SOM model higher than their own methods*

In addition, the overall satisfaction of the managers was tested according to the following hypothesis:

H_0: *Overall, the managers are neutral or dissatisfied with the SOM model*

H_1: *Overall, the managers are satisfied with the SOM model*

In order to test the hypotheses posed above, an independent samples t-test was performed on the answers. A Levene's test of equality of variance showed that the variances were unequal for content, accuracy, and ease of use. Therefore, the unequal variances assumption was used in the t-test for these factors. The results are shown in Table 9.16. Gr. in the table refers to group 1 for the managers' satisfaction with current methods, and 2 to the managers' satisfaction with the SOM model.

	Gr.	N	Mean	Std. Err. Mean	Std. Dev.	t-test for Equality of Means t	df	Sig. (2-tailed)	Mean Dif.	Std. Err. Dif.	95% Confidence Interval of the Difference Lower	Upper
	1	36	2.89	.16798	1.00791							
Content	2	36	3.82	.07943	.47660	-4.993	49.9	.000	-.928	.18582	-1.30102	-.55453
	1	36	2.94	.16400	.98400							
Accuracy	2	36	3.37	.10435	.62610	-2.172	59.4	.034	-.422	.19438	-.81113	-.03331
	1	36	2.97	.14078	.84468							
Format	2	36	3.71	.10030	.60180	-4.275	70.0	.000	-.739	.17286	-1.08364	-.39414
	1	36	2.69	.14811	.88864							
Ease of use	2	36	3.71	.05975	.35851	-6.386	46.1	.000	-1.02	.15971	-1.34140	-.69850
	1	36	3.11	.15825	.94952							
Timeliness	2	36	3.14	.14449	.86694	-.130	70.0	.897	-.028	.21429	-.45517	.39962

Table 9.16. Independent samples t-test.

The results show that the H_0 hypothesis is rejected for all of the factors except for timeliness. H_1 is thus substantiated for the factors of content, accuracy, format, and ease of use. In addition, content, format, and ease of use are significant at the 0.001 level, while accuracy is significant at the 0.05 level.

Timeliness did not differ significantly from current methods. This can be explained by the fact that this model is timely under the same assumptions that current methods are, that is, the rate at which information becomes available. This is logical as the new data can be very quickly inserted into the current model, much at the same rate as in any other model, since no retraining is required. This was also perceived by the managers.

To test the overall satisfaction hypothesis, a single sample t-test was performed, with a test value of 3 (average), i.e. to test if the managers were significantly satisfied and not neutral or dissatisfied. The results can be found in Table 9.17.

	Test Value = 3 (average) N	Mean	t	df	Sig. (2-tailed)	Mean dif.	95% confidence interval of the difference Lower	Upper
Overall Satisfaction	35	4.0286	8.613	34	0.000	1.0286	0.7859	1.2713

Table 9.17. Single sample t-test.

The results show that H_0 is rejected at the 0.001 level, and thus, H_1 is substantiated. Managers are thus significantly satisfied with the SOM model.

The results are thus very strong, under the assumptions discussed earlier.

9.6 Conclusions of the Evaluation

In the evaluation phase, a number of companies (13) were visited, and a demo of the financial benchmarking model was held. A questionnaire, based on the 5-factor Doll and Torkzadeh satisfaction instrument, was administered to the participants. A total of 36 responses (response rate 92.31%) were received.

The demographics of the population were similar to those of the state of the art survey, although the level of education was somewhat higher. Also, the targeted population, business intelligence and related tasks, was better reached. The managers' experience with basic IT tools was high, although they were much less experienced in the use of more advanced tools such as databases and decision support systems.

Compared to the respondents in the state of the art survey, the managers were significantly less satisfied with the content (p = 0.03), accuracy (p = 0.09), and ease of use (p = 0.019) of their current methods. This was likely due to that the managers that initially were less pleased with their current methods were more likely to participate in the second phase. No evidence was found to suggest that this was a result of them being influenced in their opinions by the demo. The managers' satisfaction with their current methods suggests that there is room for considerable improvements.

Roughly half of the managers reported that they often face information overload, and that complexity and turbulence in the competitive environment are high. This suggests that a complexity reducing visualization tool could provide additional benefits to managers. Support for this was also found in the results (Table 9.4).

The content of the model was highly rated by the managers. The only sub-factor of content to receive a near average rating was sufficiency. This suggests that the managers were lacking something. This was expected as the goal was to build a very general benchmarking model. Many answers to the open questions suggested that the lacking component would be qualitative analysis. Overall, however, the content of the model received the highest ratings of the factors. It was also rated the most important by the managers.

The accuracy of the model was also highly rated by the managers. Two sub-factors of accuracy, precision and completeness, received somewhat lower ratings. It was concluded that the precision of the model should be increased by providing drill-down capabilities to crisp numbers. Completeness was deemed to be problematic definition-wise (see Section 9.4.3 for discussion). Overall, accuracy also received high ratings from the managers.

The format of the model also received high ratings. The managers were quite satisfied with all sub-factors of format. This is very important support as the format and visualization are arguably the most important additions to current methods that the SOM can provide.

The ease of use of the SOM was rated fairly highly, but it was emphasized that it is not a suitable tool for daily end-user use. Instead, it was argued by the managers that the SOM would provide most value during infrequent but consistent use, such as during the strategy process, when a quick overview is required.

Timeliness was rated average, i.e. the same as for their current methods. The managers correctly perceived that the model is dependent upon the same access to data as current methods are, which means that it cannot be timelier than these. On the other hand, the timeliness was not perceived as lower than current methods.

Overall, satisfaction with the benchmarking model was very high.

Finally, six hypotheses concerning the managers' perceptions of the model were tested. The first five tested whether the managers perceived that the model was better than their own methods according to the Doll and Torkzadeh factors of information. The H_1 hypothesis (the model is better than current methods) was substantiated for the factors content $t(49.9) = -4.993$, $p = 0.000$), accuracy $t(59.4) = -2.172$, $p = 0.034$), format $t(70) = -4.275$, $p = 0.000$), and ease of use $t(46.1) = -6.386$, $p = 0.000$). Timeliness did not differ significantly from current methods. In addition, the overall satisfaction of the managers was tested against a value of 3 (neutral). The managers were significantly satisfied with the benchmarking model ($t(34) = 8.613$, $p = 0.000$).

With the support of the results of the evaluation, it can be concluded that the financial benchmarking model has been validated by the subject matter experts.

10 CONCLUSIONS AND FURTHER RESEARCH

Performing financial comparisons of companies can be very important in today's society. Such comparisons are commonly performed using benchmarking. Benchmarking is a method of objectively comparing the activities of one company to those of another, in order to find areas of improvement. Financial benchmarking, which is performance benchmarking using financial measures, is a commonly used tool in today's business world.

Today, the data required for financial benchmarking is easily available. Most companies publish annual reports, or at least abbreviated financial statements, on their homepages. The Internet has indeed become a common source of financial data. However, with the almost infinite access to information comes a problem: our capacity to process the information is not sufficient. We are in need of tools that can assist us in this task.

One such tool is data mining. Data mining is a tool for finding patterns and regularities in large amounts of data. Self-organizing maps are an example of a data-mining tool considered to be suitable for exploratory data analysis, particularly clustering and visualization problems. Self-organizing maps are two-dimensional representations of data, which group the data according to patterns or similarities in the dataset. The result is a two-dimensional topological map with light shades representing small distances, or similarities, and dark shades representing large distances, or differences.

In this thesis, a model for financial benchmarking in the international pulp and paper industry has been proposed. Firstly, a literature survey has been conducted to study the applications in which the SOM has been used. A state of the art survey has been carried out in order to determine what financial benchmarking methods are currently being used in Finnish publicly-noted companies. The managers' satisfaction with these methods was also studied, as well as the need for alternative methods.

Using the constructive research approach, a financial benchmarking model has been built and evaluated. Firstly, the key concepts financial benchmarking, knowledge discovering in databases and data mining, self-organizing maps, and financial ratios analysis were presented. Then, a database of financial information for 98 companies in the international pulp and paper industry was collected for the period 1995-2002, using the Internet as a source of information. A number of financial ratios, chosen from a previously published empirical study, were selected and calculated based on the information in the database. Then, a data-mining tool, the self-organizing map, was used to perform a financial competitor benchmarking of these companies. Finally, the created model was evaluated by subject matter experts in a face validation setting.

10.1 The Results

The preliminary survey indicated that very few advanced methods, capable of multiple ratio analysis, are used for financial benchmarking in Finnish publicly-noted companies. At the same time, the complexity of the competitive environment, combined with executives' information overload, suggest that there is a need for more advanced methods. Therefore, the SOM has been proposed as such a tool.

Using the histogram equalized financial data, a single 9×7 hexagonal map was created for the years 1995-01, and updated with data for the years 2002 and 2003. The map was used to benchmark the five largest pulp and paper companies according to net sales in 2003. The conclusion was that Kimberly-Clark's performance was considerably better than the others'. Stora Enso also displayed strong performance during 1999-2001, but performance was heavily affected by a write-down in the asset value of Consolidated Papers in 2002, as well as the poor industry conditions in 2003. The other companies in the Top 5, International Paper, Georgia-Pacific, and Weyerhaeuser, are all quite inefficient, poor performers compared to Kimberly-Clark and Stora Enso.

In the evaluation, 36 managers in business intelligence related tasks evaluated the created financial benchmarking model in a face validation setting. The evaluation consisted of a brief demo on the basics of the SOM, as well as a demonstration of benchmarking using the proposed model. Then, the managers were surveyed using a structured questionnaire based on a validated instrument for measuring end-user computing satisfaction. Five factors, content, accuracy, format, ease of use, and timeliness, as well as overall satisfaction, were evaluated. Finally, the managers' satisfaction with the model was tested using six hypotheses and statistical t-tests. All of the hypotheses except for hypothesis three (timeliness of the model) were validated. The timeliness was found to be equal to current methods. Overall, the model was thereby validated by the managers. There are two main conclusions of this work.

Firstly and most importantly, the managers considered the SOM to be a feasible tool for financial benchmarking. The primary advantage of the SOM in this case was, as expected, its visualization properties.

Secondly, a central conclusion concerning the use of the SOM was that the managers obviously felt that the greatest contribution of the SOM was in generating an overview of the competitive situation, and that the best use of the same could be made through infrequent but consistent use, such as during the strategy process. The managers did not feel that the SOM was suitable for daily use.

10.2 Future Research

With managers facing increasing amounts of information to process daily, the need for intelligent visualization tools to perform these operations is likely to increase in the future. This situation is accentuated by the exponentially increasing amount of information available through the Internet. We simply cannot cope with this information overload any longer without using intelligent tools.

Developments in information technology have increased the need for intelligent tools the feasibility of using these tools, but at the same time, have increased the feasibility of using tools like the SOM for financial data analysis. For example, the development of tag-based markup languages, such as XBRL (Extensible Business Reporting Language) will enable standardized online financial reporting (Debreceny and Gray 2001). The advantage of this is that software agents can be used to automatically retrieve specified financial information, and automatically store it in databases (e.g. Liu 1998; Nelson et al. 2000). Coupled with the reemerging interest in continuous financial reporting and auditing (Dull and Tegarden 2004), originally proposed by Vasarhelyi and Halper (1991), this will result in massive amounts of data for managers and auditors to assess. Continuous financial reporting is essentially online updating of financial reports, i.e. dynamic, real-time income statements, balance sheets, and cash flow statements. Needless to say, this will result in a huge burden on auditors, requiring an approach more similar to process state monitoring than traditional auditing. For these reasons among others, the use of data-mining tools, such as self-organizing maps, is likely to increase dramatically in the future.

In this thesis, the self-organizing map has been shown to be a feasible tool for financial benchmarking in the international pulp and paper industry. The results are easy to visualize and interpret, and provide a very practical way to compare the financial performance of different companies.

There are, however, many unexplored areas for the use of the SOM in financial benchmarking. The first and most important future area of research should be to create an interactive instantiation of the SOM model. This instantiation, a prototype, should incorporate the lessons learned in this experiment.

Firstly, it needs to be flexible so that managers can include their own ratios freely. This is because the evaluation shows that the managers were not entirely satisfied with the sufficiency of the model. A way to automatically incorporate qualitative analysis would also be important.

Drill-down capabilities and the capability to view crisp numbers would be important. The precision of the model received a relatively low rating, although not unsatisfactory, in the accuracy of the model. The managers valued the quick overview presented by the SOM, but many also suggested that it should be combined with detailed presentation of numbers, for example, in the form of tables.

The tool should not be intended for daily use, but instead during consistent but infrequent use, such as in the strategy process. Finally, the managers felt that simulation capabilities would be an important improvement. This would be especially important if the tool were to be used during strategic planning.

The prototype should be tested in a setting as described by Vesanto (2002, p.69), i.e. it should be compared to other, competing tools by fairly inexperienced end-users.

REFERENCES

Adelman, L., F. W. Rook and P. E. Lehner (1985). "User and R&D Specialist Evaluation of Decision-Support Systems." IEEE Transactions on Systems, Man, and Cybernetics **SMC-15**(3): 334-342.

Adriaans, P. and D. Zantinge (1996). "Data Mining." Addison-Wesley Longman, Harlow.

Ahmed, P. K. and M. Rafiq (1998). "Integrated benchmarking: a holistic examination of select techniques for benchmarking analysis." Benchmarking for Quality Management & Technology **5**(3): 225-242.

Alexander, D. and S. Archer (2001). "On the Myth of "Anglo-Saxon" Financial Accounting." The International Journal of Accounting **35**(4): 539-557.

Alhoniemi, E., J. Hollmén, O. Simula and J. Vesanto (1999). "Process Monitoring and Modeling using the Self-Organizing Map." Integrated Computer Aided Engineering **6**(1): 3-14.

Alter, S. (2002). "Information Systems: Foundation of E-Business." Prentice Hall, Upper Saddle River, NJ.

Ambroise, C., G. Seze, F. Badran and S. Thiria (2000). "Hierarchical Clustering of Self-Organizing Maps for Cloud Classification." Neurocomputing **30**(1): 47-52.

Anderson, T. (1996). "A Data Envelopment Analysis (DEA) Home Page." http://www.emp.pdx.edu/dea/homedea.html. (Accessed: 4.5.2004)

Athanassopoulos, A. and C. Gounaris (2001). "Assessing the technical and allocative efficiency of hospital operations in Greece and its resource allocation implications." European Journal of Operational Research **133**(2): 416-431.

Athanassopoulos, A. D. and J. A. Ballantine (1995). "Ratio and Frontier Analysis for Assessing Corporate Performance: Evidence from the Grocery Industry in the UK." Journal of the Operational Research Society **46**(4): 427-440.

Au, N., E. W. T. Ngai and T. C. E. Cheng (2002). "A critical review of end-user information system satisfaction research and a new research framework." Omega **30**(6): 451-478.

Back, B., M. Irjala, K. Sere and H. Vanharanta (1998a). "Competitive Financial Benchmarking Using Self-Organizing Maps." In Artificial Intelligence in Accounting and Auditing. Towards New Paradigms. M. A. Vasarhelyi and A. Kogan, Eds. Princeton, NJ, Marcus Wiener Publishers. **Vol IV:** 69-81.

Back, B., G. Oosterom, K. Sere and M. van Wezel (1995). "Intelligent Information Systems within Business: Bankruptcy Predictions Using Neural Networks." The 3rd European Conference on Information Systems (ECIS'95), Athens, Greece, June 1-3: 99-111.

Back, B., K. Sere and H. Vanharanta (1996). "Data Mining Accounting Numbers Using Self-organizing Maps." STeP '96 - Genes, Nets and Symbols, Vaasa, 20-23 August: 35-47.

Back, B., K. Sere and H. Vanharanta (1998b). "Managing complexity in large data bases using self-orginizing maps." Accounting Management and Information Technologies **8**(4): 191-210.

Back, B., J. Toivonen, H. Vanharanta and A. Visa (2001). "Comparing numerical data and text information from annual reports using self-orginizing maps." International Journal of Accounting Information Systems **2**(4): 249-269.

Bailay, J. E. and S. W. Pearson (1983). "Development of a tool for measuring and analyzing computer user satisfaction." Management Science **29**(5): 530-544.

Baker, R. C. and S. Talluri (1997). "A closer look at the use of data envelopment analysis for technology selection." Computers & Industrial Engineering **32**(1): 101-108.

Banker, R. D., A. Charnes and W. W. Cooper (1984). "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis." Management Science **30**(9): 1078-1092.

Baroudi, J. J., M. H. Olson and B. Ives (1986). "An empirical study of the impact of user involvement on system usage and information satisfaction." Communications of the ACM **39**(3): 232 - 238.

Baroudi, J. J. and W. J. Orlikowski (1988). "A Short-Form Measure of User Information Satisfaction: A Psychometric Evaluation and Notes on Use." Journal of Management Information Systems **4**(4): 44-59.

Becanovic, V. (2000). "Image Object Classification Using Saccadic Search, Spatio-Temporal Pattern Encoding and Self-Organization." Pattern Recognition Letters **21**(3): 253-263.

Bendell, T., L. Boulter and J. Kelly (1998). "Benchmarking for Competitive Advantage." Pitman Publishing, London.

Bertoneche, M. and R. Knight (2001). "Financial Performance." Butterworth-Heinemann, Oxford.

Bhutta, K. S. and F. Huq (1999). "Benchmarking - best practices: an integrated approach." Benchmarking: An International Journal **6**(3): 254-268.

Burrell, G. and G. Morgan (1979). "Sociological Paradigms and Organisational Analysis." Heineman, London.

Camp, R. C. (1989). "Benchmarking: The Search for Industry Best Practices that Lead to Superior Performance." American Society for Quality, Quality Press, Milwaukee, Wisconsin.

Camp, R. C. (1993). "Lär av de bästa!: Benchmarking i tio steg." Studentlitteratur, Lund.

Charnes, A., W. W. Cooper and E. L. Rhodes (1978). "Measuring the efficiency of decision making units." European Journal of Operational Research **2**(6): 429-444.

Charnes, A., J. J. Rousseau and J. H. Semple (1996). "Sensitivity and stability of efficiency classifications in data envelopment analysis." Journal of Productivity Analysis **7**(1): 5-18.

Chen, D., R. Chang and Y. Huang (2000a). "Breast Cancer Diagnosis Using Self-Organizing Map for Sonography." Ultrasound in Medicine and Biology **26**: 405-411.

Chen, L.-d., K. S. Soliman, E. Mao and M. N. Frolick (2000b). "Measuring user satisfaction with data warehouses: an exploratory study." Information & Management **37**(3): 103-110.

Chen, T.-y. (1997). "A measurement of the resource utilization efficiency of university libraries." International Journal of Production Economics **53**(1): 71-80.

Choi, F. D. S., C. A. Frost and G. K. Meek (2002). "International Accounting." Pearson Education , Inc., Upper Sadle River, N.J.

Cielen, A., L. Peeters and K. Vanhoof (2004). "Bankruptcy prediction using a data envelopment analysis." European Journal of Operational Research **154**(2): 526-532.

Collins, J. (2001). "Level 5 Leadership - The Triumpf of Humility and Fierce Resolve." Harvard Business Review **79**(1): 66-76.

Dahr, V. and R. Stein (1997). "Intelligent Decision Support Methods." Prentice-Hall, Inc, New Jersey.

Dattakumar, R. and R. Jagadeesh (2003). "A review of literature on benchmarking." Benchmarking: An International Journal **10**(3): 176-209.

Davis, F. D. (1989). "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology." MIS Quarterly **13**(3): 319-340.

Davis, F. D., R. P. Bagozzi and P. R. Warshaw (1989). "User Acceptance of Computer Technology: A Comparison of Two Theoretical Models." Management Science **35**(8): 982-1003.

de Bodt, E., M. Cottrell and M. Verleysen (2002). "Statistical tools to assess the reliability of self-organizing maps." Neural Networks **15**(8-9): 967-978.

Deboeck, G. J. (1998). "Best Practices in Data Mining Using Self-Organizing Maps." Visual Explorations in Finance with Self-Organizing Maps. G. J. Deboeck and T. Kohonen, Eds. Berlin, Springer-Verlag**:** 203-229.

Deboeck, G. J. and T. Kohonen, Eds. (1998). "Visual Explorations in Finance with Self-Organizing Maps." Springer finance. Springer-Verlag, Berlin.

Debreceny, R. and G. L. Gray (2001). "The production and use of semantically rich accounting reports on the Internet: XML and XBRL." International Journal of Accounting Information Systems **2**(1): 47-74.

DeLone, W. H. and E. R. McLean (1992). "Information Systems Success: The Quest for the Dependent Variable." Information Systems Research **3**(1): 60-95.

DeLone, W. H. and E. R. McLean (2003). "The DeLone and McLean Model of Information Systems Success: A Ten-Year Update." Journal of Management Information Systems **19**(4): 9 - 30.

Doll, W. J., A. Hendrickson and X. Deng (1998). "Using Davis's Perceived Usefulness and Ease-of-use Instruments for Decision Making: A Confirmatory and Multigroup Invariance Analysis." Decision Sciences **29**(4): 839-869.

Doll, W. J., T. S. Raghunathan, J.-S. Lim and Y. P. Gupta (1995). "A Confirmatory Factor Analysis of the User Information Satisfaction Instrument." Information Systems Research **6**(2): 177-188.

Doll, W. J. and G. Torkzadeh (1988). "The Measurement of End-User Computing Satisfaction." MIS Quarterly **12**(2): 259-274.

Doll, W. J., W. Xia and G. Torkzadeh (1994). "A confirmatory factor analysis of the end-user computing satisfaction instrument." MIS Quarterly **18**(4): 453 - 461.

Downing, C. E. (1999). "System usage behavior as a proxy for user satisfaction: an empirical investigation." Information & Management **35**(4): 203-216.

Doyle, J. R. (1995). "Multiattribute Choice for the Lazy Decision Maker: Let the Alternatives Decide!" Organizational Behavior and Human Decision Processes **62**(1): 87-100.

Dull, R. B. and D. P. Tegarden (2004). "Using control charts to monitor financial reporting of public companies." International Journal of Accounting Information Systems **5**(2): 109-127.

Eklund, T., B. Back, H. Vanharanta and A. Visa (2001). "Benchmarking Global Pulp and Paper Companies using Self-Organizing Maps." Paperi ja Puu, Pulp and Timber Magazine **83**(4): 304-316.

Eklund, T., B. Back, H. Vanharanta and A. Visa (2002). "Assessing the Feasibility of Self-Organizing Maps for Data Mining Financial Information." The Xth European Conference on Information Systems (ECIS 2002), Gdansk, Poland, June 6-8: 528-537.

El-Gazzar, S. M., P. M. Finn and R. Jacob (1999). "An Empirical Investigation of Multinational Firms' Compliance with International Accounting Standards." The International Journal of Accounting **34**(2): 239-248.

Emrouznejad, A. (2003). "Data Envelopment Analysis Home Page." http://www.DEAzone.com. (Accessed: 4.5.2004)

Etezadi-Amoli, J. and A. F. Farhoomand (1996). "A structural model of end user computing satisfaction and user performance." Information & Management **30**(2): 65-73.

Fayyad, U., G. Piatetsky-Shapiro and P. Smyth (1996). "The KDD Process for Extracting Useful Knowledge from Volumes of Data." Communications of the ACM **39**(11): 27-34.

Feroz, E. H., S. Kim and R. Raab (2003). "Financial statement analysis: a data envelopment analysis approach." Journal of the Operational Research Society **54**(1): 48-58.

Finnish Forest Industries Federation (2000). "Major Paper and Paperboard Producers 1999." http://www.forestindustries.fi/en/tilastot/suuripk.htm. (Accessed: 6.9.2000)

Finnish Forest Industries Federation (2004). "Main Global Forest Industry Companies Turnover in 2003, bill. EUR." Finnish Forest Industries Federation. http://english.forestindustries.fi/figures/. (Accessed: 13.05.2004)

Flexer, A. (2001). "On the use of self-organizing maps for clustering and visualization." Intelligent Data Analysis **5**(5): 373-384.

Foster, G. (1978). "Financial Statement Analysis." Prentice-Hall, Inc., Englewood Cliffs, N.J.

Galletta, D. F. and A. L. Lederer (1989). "Some Cautions on the Measurement of User Information Satisfaction." Decision Sciences **20**(3): 419-438.

Galliers, R. D. (1992). "Choosing Information Systems Research Approaches." Information System Research - Issues, methods and practical guidelines. R. D. Galliers, Ed. Oxford, Blackwell Scientific Publications**:** 144-162.

Galliers, R. D. and F. F. Land (1987). "Choosing appropriate information systems research methodologies." Communications of the ACM **30**(11): 901-902.

Gelderman, M. (1998). "The relation between user satisfaction, usage of information systems and performance." Information & Management **34**(1): 11-18.

Goles, T. and R. Hirschheim (2000). "The paradigm is dead, the paradigm is dead...long live the paradigm: the legacy of Burrell and Morgan." Omega **28**(3): 249-268.

González, E. and A. Álvarez (2001). "From efficiency measurement to efficiency improvement: The choice of a relevant benchmark." European Journal of Operational Research **33**(3): 512-520.

Goodhue, D. L. (1995). "Understanding User Evaluations of Information Systems." Management Science **41**(12): 1827-1844.

Goodhue, D. L., B. D. Klein and S. T. March (2000). "User evaluations of IS as surrogates for objective performance." Information & Management **38**(2): 87-101.

Gordon, D. and G. Geiger (1999). "Strategic Management of an Electronic Patient Record Project Using the Balanced Scorecard." Journal of Healthcare Information Management **13**(3): 113-123.

Guilding, C. (1999). "Competitor-focused accounting: an exploratory note." Accounting, Organizations and Society **24**(7): 583-595.

Guiver, J. P. and C. C. Klimasauskas (1991). "Applying Neural Networks, Part IV: Improving Performance." PC/AI Magazine **5**(4): 34-41.

Gustafsson, L. (1992). "Bäst i klassen: Benchmarking för högre effektivitet." Sveriges Verkstadsindustrier, Uppsala.

Harris, M. S. and K. A. Muller III (1999). "The market valuation of IAS versus US-GAAP accounting measures using Form 20-F reconciliations1." Journal of Accounting and Economics **26**(1-3): 285–312.

Haykin, S. (1999). "Neural Networks - A Comprehensive Foundation." Prentice Hall International, Inc., Upper Saddle River, N.J.

Hebb, D. O. (1949). "The Organization of Behavior." Wiley, New York.
Hevner, A. R., S. T. March, J. Park and S. Ram (2004). "Design Science in Information Systems Research." MIS Quarterly **28**(1): 75-105.
Hill, R. (1998). "What Sample Size is "Enough" in Internet Survey Research?" Interpersonal Computing and Technology: An Electronic Journal for the 21st Century **6**(3-4). http://www.emoderators.com/ipct-j/1998/n3-4/hill.html. (Accessed: 21.04.02)
Himberg, J., J. Ahola, J. Vesanto and O. Simula (2001). "The Self-Organizing Map as a Tool in Knowledge Engineering." Pattern Recognition in Soft Computing. N. R. Pal, Ed. River Edge, World Scientific. **2:** 38-65.
Hirschheim, R. and H. K. Klein (1989). "Four paradigms of information systems development." Communications of the ACM **32**(10): 1199-1216.
Hollmén, J. (2000). "User profiling and classification for fraud detection in mobile communications networks." Helsinki University of Technology, Espoo.
Hollmén, J., V. Tresp and O. Simula (1999). "A self-organizing map for clustering probabilistic models." The Ninth International Conference on Artificial Neural Networks (ICANN'99), Edinburgh, UK, 7-10 September, IEE: 946–951.
Homburg, C. (2001). "Using data envelopment analysis to benchmark activities." International Journal of Production Economics **73**(1): 51-58.
Honkela, T., S. Kaski, K. Lagus and T. Kohonen (1997). "WEBSOM - Self-Organizing Maps of Document Collections." WSOM'97: Workshop on Self-Organizing Maps, Espoo, June, Helsinki University of Technology: 310-315.
Igbaria, M. and M. Tan (1997). "The consequences of information technology acceptance on subsequent individual performance." Information & Management **32**(3): 113-121.
Iivari, J. (1991). "A Paradigmatic Analysis of Contemporary Schools of IS Development." European Journal of Information Systems **1**(4): 249-272.
Iivari, J., R. Hirschheim and H. K. Klein (1998). "A Paradigmatic Analysis Contrasting Information Systems Development Approaches and Methodologies." Information Systems Research **9**(2): 164-193.
Ives, B., M. H. Olson and J. J. Baroudi (1983). "The measurement of user information satisfaction." Communications of the ACM **26**(10): 785-793.
Järvinen, P. (2001). "On Research Methods." Opinpaja OY, Tampere.
Kangas, J. (1994). "On the Analysis of Pattern Sequences by Self-Organizing Maps." Helsinki University of Technology, Helsinki.
Kaplan, R. S. and D. P. Norton (1992). "The Balanced Scorecard **-** Measures that Drive Performance." Harvard Business Review **70**(1): 71-80.
Kaplan, R. S. and D. P. Norton (1993). "Putting the Balanced Scorecard to Work." Harvard Business Review **71**(6): 134-147.
Karlöf, B. (1997). "Benchmarking i verkligheten: De goda förebildernas inspiration till lärande och affärsutveckling." Werner Söderström, Borgå.

Karlöf, B. and S. Östblom (1994). "Benchmarking: Vägvisare till mästerskap i produktivitet och kvalitet." Werner Söderström, Borgå.

Kasanen, E., K. Lukka and A. Siitonen (1993). "The Constructive Approach in Management Accounting Research." Journal of Management Accounting Research **5**: 243-264.

Kaski, S. (1997). "Data Exploration Using Self-Organizing Maps." Acta Polytechnica Scandinavia, Mathematics, Computing and Management in Engineering Series, No. 82, Helsinki University of Technology, Helsinki.

Kaski, S., J. Kangas and T. Kohonen (1998). "Bibliography of Self-Organizing Map (SOM) Papers 1981-1997." Neural Computing Surveys. http://www.icsi.berkeley.edu/~jagota/NCS/. (Accessed: July 24th 2000)

Kaski, S. and T. Kohonen (1996). "Exploratory Data Analysis by the Self-Organizing Map: Structures of Welfare and Poverty in the World." Neural Networks in Financial Engineering. P. N. Apostolos, Y. A.-M. Refenes, J. Moody and A. Weigend, Eds. Singapore, World Scientific**:** 498-507.

Kaski, S. and K. Lagus (1996). "Comparing Self-Organizing Maps." International Conference on Artificial Neural Networks (ICANN'96), Springer: 809-814.

Kiang, M. Y. and A. Kumar (2001). "An Evaluation of Self-Organizing Map Networks as a Robust Alternative to Factor Analysis in Data Mining Applications." Information Systems Research **12**(2): 34-41.

Kiviluoto, K. (1998). "Predicting bankruptcies with the self-organizing map." Neurocomputing **21**(1-3): 191-201.

Kloptchenko, A. (2003). "Text Mining Based on the Prototype Matching Method." TUCS Dissertations No 47, Turku Centre for Computer Science, Turku.

Kohonen, T. (1972). "Correlation Matrix Memories." IEEE Transactions on Computers **C-21**: 353-359.

Kohonen, T. (1989). "Self-Organization and Associative Memory." Springer-Verlag, Heidelberg.

Kohonen, T. (1990). "The Self-Organizing Map." Proceedings of the IEEE **78**(9): 1464-1480.

Kohonen, T. (1997). "Self-Organizing Maps." Springer-Verlag, Leipzig, Germany.

Kohonen, T. (2001). "Self-Organizing Maps." Springer-Verlag, Berlin.

Kohonen, T., J. Hynninen, J. Kangas and J. Laaksonen (1996). "SOM_PAK: The Self-Organizing Map Program Package." No.,Helsinki University of Technology Helsinki.

Kohonen, T., S. Kaski, K. Lagus, J. Salojärvi, J. Honkela, V. Paatero and A. Saarela (2000). "Self organization of a massive document collection." IEEE Transactions on Neural Networks **11**(3): 574-585.

Lagus, K. (2000). "Text Mining with WEBSOM." Helsinki University of Technology, Espoo, Finland.

Lainez, J. A. and S. Callao (2000). "The Effect of Accounting Diversity on International Financial Analysis: Empirical Evidence." The International Journal of Accounting **35**(1): 65-83.

Länsiluoto, A., B. Back, H. Vanharanta and A. Visa (2002). "Multivariable Business Cycle Analysis with Self-Organizing Maps - Are the Cycles Similar?" European Conference on Accounting Information System (ECAIS 2002), Copenhagen, 23-24 April:

Lehtinen, J. (1996). "Financial Ratios in an International Comparison: Validity and Reliability." Universitas Wasaensis, Vasa.

Lev, B. (1974). "Financial statement analysis: a new approach." Prentice-Hall, Inc., Englewood Cliffs, N.J.

Li, E. Y. (1997). "Perceived importance of information system success factors: A meta analysis of group differences." Information & Management **32**(1): 15-28.

Lincoln, Y. S. and E. G. Guba (2000). "Paradigmatic Controversies, Contradictions, and Emerging Confluences." Handbook of Qualitative Research, Second Edition. N. K. Denzin and Y. S. Lincoln, Eds. Thousand Oaks, CA, Sage Publications**:** 163-188.

Liu, S. (1998). "Business Environment Scanner for Senior Managers: Towards Active Executive Support with Intelligent Agents." Expert Systems with Applications **15**(2): 111-121.

Mangiameli, P., S. K. Chen and D. West (1996). "A comparison of SOM neural network and hierarchical clustering methods." European Journal of Operational Research **93**(2): 402-417.

March, S. T. and G. F. Smith (1995). "Design and natural science research on information technology." Decision Support Systems **15**(4): 251-266.

Martín-del-Brío, B. and C. Serrano-Cinca (1993). "Self-organizing Neural Networks for the Analysis and Representation of Data: Some Financial Cases." Neural Computing and Applications **1**(2): 193-206.

McCulloch, W. and W. Pitts (1943). "A Logical Calculus of the Ideas Immanent in Nervous Activity." Bulletin of Mathematical Biophysics **5**: 115-133.

McNair, C. J. and K. H. J. Liebfried (1992). "Benchmarking - A Tool for Continuous Improvement." John Wiley and Sons, Inc., New York, N.Y.

Melone, N. P. (1990). "A Theoretical Assessment of the User-Satisfaction Construct in Information Systems Research." Management Science **36**(1): 76-91.

Miller, G. A. (1956). "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information." Psychological Review **63**(2): 81-97.

Mingers, J. (2001). "Combining IS Research Methods: Towards a Pluralist Methodology." Information Systems Research **12**(3): 240-259.

Moxon, B. (1996). "Defining Data Mining." DBMS Online, August 1996. http://www.dbmsmag.com/9608d53.html. (Accessed: 26.6.2000)

Murphy, K. E. and S. J. Simon (2002). "Intangible benefits valuation in ERP projects." Information Systems Journal **12**(4): 301-320.

Muylle, S., R. Moenaert and M. Despontin (2004). "The conceptualization and empirical validation of web site user satisfaction." Information & Management **41**(5): 543-560.

Nelson, K. M., A. Kogan, R. P. Srivastava, M. A. Vasarhelyi and H. Lu (2000). "Virtual auditing agents: the EDGAR Agent challenge." Decision Support Systems **28**(3): 241-253.

Nobes, C. (2002). "International classification of financial reporting." Comparative International Accounting. C. Nobes and R. Parker, Eds. Edinborough Gate, Pearson Education Limited**:** 52-71.

Nobes, C. (2003). "On the myth of "Anglo-Saxon" financial accounting: a comment." The International Journal of Accounting **38**(1): 95-104.

Nobes, C. and R. Parker (1991). "Comparative International Accounting." Prentice Hall International, Cambridge.

Nobes, C. and R. Parker, Eds. (2000). "Comparative International Accounting. Pearson Education Limited, Essex.

Nobes, C. and R. Parker (2002). "Comparative International Accounting." Pearson Education Limited, Edinburgh Gate, Essex.

Nunamaker, J. F. J., C. Minder and T. D. M. Purdin (1991). "Systems Developement in Information Systems Research." Journal of Management Information Systems **7**(3): 89-106.

Oja, E., S. Kaski and T. Kohonen (2003). "Bibliography of Self-Organizing Map (SOM) Papers: 1998-2001 Addendum." Neural Computing Surveys **3**: 1-156.

Oja, E. and K. Kiviluoto (1999). "Data Analysis, Visualization, and Hidden Factor Discovery by Unsupervised Learning." International Joint Conference on Neural Networks (IJCNN'99), Washington, D.C., July 10-16, IEEE Neural Networks Council: Piscataway:

Olmsted, D. D. (1998). "History and Principles of Neural Networks." http://www.pilotsw.com/datamining/dmindex.html. (Accessed: 26.6.2000)

Ong, C.-S. and J.-Y. Lai (2004). "Developing an instrument for measuring user satisfaction with knowledge management systems." 37th Hawaii International Conference on System Sciences (HICSS-37), Big Island, Hawaii, IEEE:

Page, E. H., B. S. Canova and J. A. Tufarolo (1997). "A Case Study of Verification, Validation, and Accreditation for Advanced Distributed Simulation." ACM Transactions on Modeling and Computer Simulation **7**(3): 393-424.

Pinsonneault, A. and K. L. Kraemer (1993). "Survey Research Methodology in Management Information Systems: An Assessment." Journal of Management Information Systems **10**(2): 75-106.

Plumlee, M. A. (2003). "The Effect of Information Complexity on Analysts' Use of That Information." The Accounting Review **78**(1): 275-296.

Post, T. and J. Spronk (1999). "Performance benchmarking using interactive data envelopment analysis." European Journal of Operational Research **155**(3): 472-487.

Rai, A., S. S. Lang and R. B. Welker (2002). "Assessing the Validity of IS Success Models: An Empirical Test and Theoretical Analysis." Information Systems Research **13**(1): 50-69.

Reider, R. (2000). "Benchmarking Strategies: A Tool For Profit Improvement." John Wiley & Sons, Inc., New York.

Rhiannon, J., C. Jewitt, L. Galasso and G. Fortemps (2001). "Consolidation Changes the Shape of the Top 150." Pulp and Paper International **43**(9): 31-41.

Rickards, R. C. (2003). "Setting benchmarks and evaluating balanced scorecards with data envelopment analysis." Benchmarking: An International Journal **10**(3): 226-245.

Roscoe, J. T. (1975). "Fundamental Research Statistics for the Behavioural Sciences." Holt Rinehart & Winston, New York.

Rumelhart, D. E. and J. L. McClelland, Eds. (1986). "Parallel Distributed Processing : Explorations in the Microstructure of Cognition. MIT Press, Cambridge, Ma.

Saarinen, T. (1996). "An expanded instrument for evaluating information system success." Information & Management **31**(2): 103-118.

Salmi, T. and T. Martikainen (1994). "A Review of the Theoretical and Empirical Basis of Financial Ratio Analysis." The Finnish Journal of Business Economics(4): 426-448.

Salonen, H. and H. Vanharanta (1990a). "Financial Analysis World Pulp and Paper Companies 1985-1989, Europe." Ekono, Espoo.

Salonen, H. and H. Vanharanta (1990b). "Financial Analysis World Pulp and Paper Companies 1985-1989, Nordic Countries." Ekono, Espoo.

Salonen, H. and H. Vanharanta (1990c). "Financial Analysis World Pulp and Paper Companies 1985-1989, North America." Ekono, Espoo.

Sarle, W. S. (2001). "Neural Network FAQ, monthly posting to the Usenet newsgroup comp.ai.neural-nets." ftp://ftp.sas.com/pub/neural/FAQ.html. (Accessed: August 2001)

Seddon, P. B. (1997). "A respecification and extension of the DeLone and McLean model of IS success." Information Systems Research **8**(3): 240-253.

Seddon, P. B. and S.-K. Yip (1992). "Empirical Evaluation of User Information Satisfaction (UIS) Measures for Use with General Ledger Accounting Software." Journal of Information Systems **6**(1): 75-92.

Serrano-Cinca, C. (1996). "Self organizing neural networks for financial diagnosis." Decision Support Systems **17**(3): 227-238.

Shanker, M., M. Hu and M. Hung (1996). "Effect of Data Standardization on Neural Network Training." Omega **24**(4): 385-397.

Simon, H. A. (1969). "The Sciences of the Artificial." MIT Press, Cambridge.

Simula, O., E. Alhoniemi, J. Hollmén and J. Vesanto (1996). "Monitoring and Modeling of Complex Processes using Hierarchical Self-Organizing Maps." (IEEE) International Symposium on Circuits and Systems (ICAIS '96), May: 73-76.

Simula, O., P. Vasara, J. Vesanto and R.-R. Helminen (1999a). "The Self-Organizing Map in Industry Analysis." Industrial Application of Neural Networks. L. C. Lain and V. R. Vemuri, Eds. London, CRC Press**:** 87-112.

Simula, O., J. Vesanto, E. Alhoniemi and J. Hollmén (1999b). "Analysis and Modeling of Complex Systems Using the Self-Organizing Map." Neuro-Fuzzy Techniques for Intelligent Information Systems. N. Kasabov and R. Kozma, Eds. Heidelberg, Springer Verlag**:** 3-22.

Siponen, M., J. Vesanto, O. Simula and P. Vasara (2001). "An approach to automated interpretation of SOM." Advances in Self-Organizing Maps. N. Allinson, H. Yin, L. Allinson and J. Slack, Eds. Berlin, Springer**:** 89-94.

Smith, P. C. (1990). "Data envelopment analysis applied to financial statements." Omega **18**(2): 131-138.

Somers, T. M., K. Nelson and J. Karimi (2003). "Confirmatory Factor Analysis of the End-User Computing Satisfaction Instrument: Replication within an ERP Domain." Decision Sciences **34**(3): 595-621.

Speidell, L. S. and V. B. Bavishi (1992). "GAAP Arbitrage: Valuation Opportunities in International Accounting Standards." Financial Analysts Journal **48**(6): 58-66.

Spiegler, I. (2003). "Technology and knowledge: bridging a ‘‘generating’’ gap." Information & Management **40**(6): 533–539.

Steering Committee for the Review of Commonwealth/State Service Provision (1997). "Data Envelopment Analysis: A technique for measuring the efficiency of government service delivery." No.,AGPS Canberra.

Street, D. L. (2002). "GAAP 2001 - Benchmarking national accounting standards against IAS: summary of results." Journal of International Accounting, Auditing & Taxation **11**(1): 77–90.

Street, D. L. and S. M. Bryant (2000). "Disclosure Level and Compliance with IASs: A Comparison of Companies With and Without U.S. Listings and Filings." International Journal of Accounting **35**(3): 305-329.

Street, D. L., S. J. Gray and S. M. Bryant (1999). "Acceptance and Observance of International Accounting Standards: An Empirical Study of Companies Claiming to Comply with IASs." The International Journal of Accounting **34**(1): 11-48.

Street, D. L., N. B. Nichols and S. J. Gray (2000). "Assessing the Acceptability of International Accounting Standards in the US: An Empirical Study of the Materiality of US GAAP Reconciliations by Non-US Companies Complying with IASC Standards." International Journal of Accounting **35**(1): 27-63.

Tan, R. G. H., J. van den Berg and W.-H. van den Bergh (2002). "Credit Rating Classification Using Self-Orginizing Maps." Neural Networks in Business: Techniques and Applications. K. Smith and J. Gupta, Eds. Hershey, Idea Group Publishing**:** 140-153.

Tavares, G. (2002). "A Bibliography of Data Envelopment Analysis (1978-2001)." Rutcor Research Report No. RRR 01-02,Rutgers Center for Operations Research, Rutgers University Piscataway, New Jersey.

Taylor, M. E. and R. A. Jones (1999). "The Use of International Accounting Standards Terminology, a Survey of IAS Compliance Disclosure." The International Journal of Accounting **34**(4): 557-570.

Tryba, V. and K. Goser (1991). "Self-Organizing Feature Maps for Process Control in Chemistry." Artificial Neural Networks. T. Kohonen, K. Mäkisara, O. Simula and J. Kangas, Eds. Amsterdam, North-Holland. **1:** 847-852.

Tull, D. S. and D. I. Hawkins (1987). "Marketing Research - Measurement and Method." Macmillan Publishing Company, New York.

Ultsch, A. (1993). "Self organized feature planes for monitoring and knowledge acquisition of a chemical process." The International Conference on Artificial Neural Networks, Springer-Verlag: London: 864-867.

Vasarhelyi, M. A. and F. B. Halper (1991). "The continuous audit of online systems." Auditing, A Journal of Practice and Theory **19**(1): 110-125.

Vesanto, J. (2002). "Data Exploration Process Based on the Self-Organizing Map." Helsinki University of Technology, Espoo.

Vesanto, J. and E. Alhoniemi (2000). "Clustering of the Self-Organizing Map." IEEE Transactions on Neural Networks **11**(3): 586-600.

Vesanto, J., J. Himberg, E. Alhoniemi and J. Parhankangas (2000). "SOM Toolbox for Matlab 5." Technical Report No. A57,Helsinki University of Technology Espoo.

Visa, A., J. Toivonen, B. Back and H. Vanharanta (2000). "A New Methodology for Knowledge Retrieval from Text Documents." TOOLMET2000 Symposium - Tool Environments and Development Methods for Intelligent Systems: 147-151.

Visa, A., J. Toivonen, B. Back and H. Vanharanta (2002). "Contents Matching Defined by Prototypes: Methodology Verification with Books of the Bible." Journal of Management Information Systems **18**(4): 87-100.

Viswanathan, M., S. Sudman and M. Johnson (2004). "Maximum versus meaningful discrimination in scale response: Implications for validity of measurement of consumer perceptions about products." Journal of Business Research **57**(2): 108-124.

Wade, N. J. (2001). "The Bains of Psychology." Perception **30**(7): 777-783.

Wang, S. (2001). "Cluster Analysis Using a Validated Self-Organizing Method: Cases of Problem Identification." International Journal of Intelligent Systems in Accounting, Finance and Management **10**(2): 127-138.

Wang, S. and H. Wang (2002). "Knowledge Discovery Through Self-Organizing Maps: Data Visualization and Query Processing." Knowledge and Information Systems **4**(1): 31-45.

Wasserman, P. D. (1989). "Neural Computing - Theory and Practice." Van Nostrand Reinhold, New York.

Whittington, M. (2000). "Problems in Comparing Financial Performance Across International Boundaries: A Case Study Approach." International Journal of Accounting **35**(3): 399-413.

Wu, M., M. Fuller and R. Wilkinson (2001). "Using Clustering and Classification Approaches in Interactive Retrieval." Information Processing and Management **37**(3): 459-484.

Yun, Y. B., H. Nakayama and T. Tanino (2004). "A generalized model for data envelopment analysis." European Journal of Operational Research **157**(1): 87-105.

Zairi, M. (1994). "Benchmarking: The Best Tool for Measuring Competitiveness." Benchmarking for Quality Management and Technology **1**(1): 11-24.

Zhu, J. (1996). "Data Envelopment Analysis with Preference Structure." Journal of the Operational Research Society **47**(1): 136-150.

PART II
Original Research Papers

Research Paper 1

Eklund, T., B. Back, H. Vanharanta and A. Visa (2004). *Financial Benchmarking Tools in Finnish Companies - A State of the Art Survey*. TUCS Technical Report No. 618, Turku, Turku Centre for Computer Science.

Research Paper 2

Eklund, T., B. Back, H. Vanharanta and A. Visa (2003). Financial Benchmarking Using Self-Organizing Maps – Studying the International Pulp and Paper Industry. Data Mining - Opportunities and Challenges. J. Wang. Hershey, PA, Idea Group Publishing: 323-349. Preprint version.

Research Paper 3

Eklund, T., B. Back, H. Vanharanta and A. Visa (2003). "Using the self-organizing map as a visualization tool in financial benchmarking." Information Visualization 2(3): 171-181.

Research Paper 4

Länsiluoto, A., T. Eklund, B. Back, H. Vanharanta and A. Visa (2004). "Industry Specific Cycles and Companies' Financial Performance - Comparison with Self-Organizing Maps." Benchmarking: An International Journal 11(4): pp. 267-286.

Research Paper 5

Kloptchenko, A., T. Eklund, B. Back, J. Karlsson, H. Vanharanta and A. Visa (2004). "Combining data and text mining techniques for analyzing financial reports." International Journal of Intelligent Systems in Accounting, Finance and Management 12(1): 29-41.

Research Paper 6

Magnusson, C., A. Arppe, T. Eklund, B. Back, H. Vanharanta and A. Visa (200x). "The language of quarterly reports as an indicator of change in the company's financial status." Accepted for publication in *Information & Management*.

Research Paper 7

Eklund, T., B. Back, H. Vanharanta, and A. Visa (200x). “Evaluating a SOM-Based Financial Benchmarking Tool”. Submitted to the *International Journal of Accounting Information Systems* (in review process).

Turku Centre for Computer Science
TUCS Dissertations

19. **Vesa Torvinen**, Construction and Evaluation of the Labour Game Method
20. **Jorma Boberg**, Cluster Analysis. A Mathematical Approach with Applications to Protein Structures
21. **Leonid Mikhajlov**, Software Reuse Mechanisms and Techniques: Safety Versus Flexibility
22. **Timo Kaukoranta**, Iterative and Hierarchical Methods for Codebook Generation in Vector Quantization
23. **Gábor Magyar**, On Solution Approaches for Some Industrially Motivated Combinatorial Optimization Problems
24. **Linas Laibinis**, Mechanised Formal Reasoning About Modular Programs
25. **Shuhua Liu**, Improving Executive Support in Strategic Scanning with Software Agent Systems
26. **Jaakko Järvi**, New Techniques in Generic Programming : C++ is More Intentional than Intended
27. **Jan-Christian Lehtinen**, Reproducing Kernel Splines in the Analysis of Medical Data
28. **Martin Büchi**, Safe Language Mechanisms for Modularization and Concurrency
29. **Elena Troubitsyna**, Stepwise Development of Dependable Systems
30. **Janne Näppi**, Computer-Assisted Diagnosis of Breast Calcifications
31. **Jianming Liang**, Dynamic Chest Images Analysis
32. **Tiberiu Seceleanu**, Systematic Design of Synchronous Digital Circuits
33. **Tero Aittokallio**, Characterization and Modelling of the Cardiorespiratory System in Sleep-disordered Breathing
34. **Ivan Porres**, Modeling and Analyzing Software Behavior in UML
35. **Mauno Rönkkö**, Stepwise Development of Hybrid Systems
36. **Jouni Smed**, Production Planning in Printed Circuit Board Assembly
37. **Vesa Halava**, The Post Correspondence Problem for Marked Morphisms
38. **Ion Petre**, Commutation Problems on Sets of Words and Formal Power Series
39. **Vladimir Kvassov**, Information Technology and the Productivity of Managerial Work
40. **Franck Tétard**, Managers, Fragmentation of Working Time, and Information Systems
41. **Jan Manuch**, Defect Theorems and Infinite Words
42. **Kalle Ranto**, Z_4-Goethals Codes, Decoding and Designs
43. **Arto Lepistö**, On Relations between Local and Global Periodicity
44. **Mika Hirvensalo**, Studies on Boolean Functions Related to Quantum Computing
45. **Pentti Virtanen**, Measuring and Improving Component-Based Software Development
46. **Adekunle Okunoye**, Knowledge Management and Global Diversity - A Framework to Support Organisations in Developing Countries
47. **Antonina Kloptchenko**, Text Mining Based on the Prototype Matching Method
48. **Juha Kivijärvi**, Optimization Methods for Clustering
49. **Rimvydas Rukšėnas**, Formal Development of Concurrent Components
50. **Dirk Nowotka**, Periodicity and Unbordered Factors of Words
51. **Attila Gyenesei**, Discovering Frequent Fuzzy Patterns in Relations of Quantitative Attributes
52. **Petteri Kaitovaara**, Packaging of IT Services – Conceptual and Empirical Studies
53. **Petri Rosendahl**, Niho Type Cross-Correlation Functions and Related Equations
54. **Péter Majlender**, A Normative Approach to Possibility Theory and Soft Decision Support
55. **Seppo Virtanen**, A Framework for Rapid Design and Evaluation of Protocol Processors
56. **Tomas Eklund**, The Self-Organizing Map in Financial Benchmarking

Turku Centre *for* Computer Science

Lemminkäisenkatu 14 A, 20520 Turku, Finland | www.tucs.fi

University of Turku

- Department of Information Technology
- Department of Mathematics

Åbo Akademi University

- Department of Computer Science
- Institute for Advanced Management Systems Research

Turku School of Economics and Business Administration

- Institute of Information Systems Sciences

ISBN 952-12-1436-8
ISSN 1239-1883