TUCS

Hongyan Liu

On Advancing Business Intelligence in the Electricity Retail Market

Turku Centre *for* Computer Science

TUCS Dissertations
No 182, September 2014

On Advancing Business Intelligence in the Electricity Retail Market

Hongyan Liu

To be presented, with the permission of the Department of Information Technologies of Åbo Akademi University, for public criticism in Auditorium Gamma, at the ICT building, Turku, Finland, on September 30th, 2014, at 12 noon.

Åbo Akademi University
Department of Information Technologies
Turku Centre for Computer Science
Data Mining and Knowledge Management Research Unit
Joukahaisenkatu 3-5 A
20520 Turku
Finland

2014

Supervised by

Professor Barbro Back
Department of Information Technologies
Åbo Akademi University
Turku, Finland

Docent Tomas Eklund
Department of Information Technologies
Åbo Akademi University
Turku, Finland

Reviewed by

Akademischer Oberrat Dr. Henning Baars
Institute of Business Administration
Department VII
Stuttgart University
Stuttgart, Germany

Professor Jarmo Partanen
LUT Energy, Faculty of Technology
Lappeenranta University of Technology
Lappeenranta, Finland

Opponent

Akademischer Oberrat Dr. Henning Baars
Institute of Business Administration
Department VII
Stuttgart University
Stuttgart, Germany

ISSN 1239-1883

Turku 2014

Abstract

In recent decades, business intelligence (BI) has gained momentum in real-world practice. At the same time, business intelligence has evolved as an important research subject of Information Systems (IS) within the decision support domain. Today's growing competitive pressure in business has led to increased needs for real-time analytics, i.e., so called real-time BI or operational BI. This is especially true with respect to the electricity production, transmission, distribution, and retail business since the law of physics determines that electricity as a commodity is nearly impossible to be stored economically, and therefore demand-supply needs to be constantly in balance.

The current power sector is subject to complex changes, innovation opportunities, and technical and regulatory constraints. These range from low carbon transition, renewable energy sources (RES) development, market design to new technologies (e.g., smart metering, smart grids, electric vehicles, etc.), and new independent power producers (e.g., commercial buildings or households with rooftop solar panel installments, a.k.a. Distributed Generation). Among them, the ongoing deployment of Advanced Metering Infrastructure (AMI) has profound impacts on the electricity retail market.

From the view point of BI research, the AMI is enabling real-time or near real-time analytics in the electricity retail business. Following Design Science Research (DSR) paradigm in the IS field, this research presents four aspects of BI for efficient pricing in a competitive electricity retail market:

(i) visual data-mining based descriptive analytics, namely electricity consumption profiling, for pricing decision-making support;
(ii) real-time BI enterprise architecture for enhancing management's capacity on real-time decision-making;
(iii) prescriptive analytics through agent-based modeling for price-responsive demand simulation;

(iv) visual data-mining application for electricity distribution benchmarking.

Even though this study is from the perspective of the European electricity industry, particularly focused on Finland and Estonia, the BI approaches investigated can:

(i) provide managerial implications to support the utility's pricing decision-making;
(ii) add empirical knowledge to the landscape of BI research;
(iii) be transferred to a wide body of practice in the power sector and BI research community.

Sammandrag

De senaste årtiondena har användningen av affärsintelligens (Eng. Business Intelligence, BI) i organisationer tagit fart. Samtidigt har BI utvecklats som ett viktigt forskningsämne i informationssystem inom området beslutsstöd. Dagens ökande konkurrenstryck i näringslivet har lett till ökade behov av realtidsanalyser, det vill säga så kallade realtids-BI eller operativa BI-system Det här gäller särskilt inom elproduktion, elöverföring, eldistribution och detaljhandel av el, eftersom fysikens lagar ger vid handen att elektricitet, som en vara är nästan omöjlig att lagra på ett ekonomiskt sätt. Därför måste utbud och efterfrågan konstant vara i balans.

Dagens kraftsektor är föremål för komplexa förändringar, innovationsmöjligheter, tekniska begränsningar och tillsynskrav. Dessa sträcker sig från övergången till lägre koldioxidutsläppsnivåer, utveckling av förnybara energikällor och skapandet av (nya) marknadsstrukturer till ny teknik (t.ex. smarta mätare, smarta nät, elbilar, etc) och nya oberoende kraftproducenter (t.ex. kommersiella byggnader eller hushåll med solpaneler, så kallad distribuerad produktion). Som ett exempel på detta kan nämnas hur den pågående utbyggnaden av en avancerad infrastruktur för mätaravläsning (AIM) har en betydande inverkan på elhandelsmarknaden.

Ur BI-forskningens synvinkel möjliggör AIM att i eller nära nog realtid analysera elhandelsverksamhet. Med hjälp av den konstruktiva forskningsansatsen presenterar denna forskning fyra aspekter av BI för att åstadkomma en effektiv prissättning på en konkurrensutsatt elhandelsmarknad:

(i) en beskrivande analys baserad på visuell datautvinning, i form av profilering av elförbrukning för att stöda prissättningsbeslut,
(ii) en företagsarkitektur för realtids-BI för att stöda ledningen för bättre beslut i realtid,

(iii) en preskriptiv analys av priskänslig efterfrågan med hjälp av agentbaserad simulering och
(iv) en visuell datautvinningstillämpning för benchmarking inom eldistribution.

Även om denna studie utgår från den europeiska elindustrin, särskilt inriktat på Finland och Estland kan de undersökta strategierna för BI:

(i) ge slutsatser till stöd för energibolagens prissättningsbeslut,
(ii) tillfoga empirisk kunskap för BI-forskningen och
(iii) överföras och tillämpas brett som praxis inom kraftsektorn och BI-forskarsamfundet.

Acknowledgments

I share the view that ‘life is a collection of experiences’. Therefore, spending 4-5 years to learn how to do scholarly research became one of my life experience collections. For this reason, I would like to thank Prof. Barbro Back for such an opportunity. The experience has been stretched a bit longer than it was originally planned, though.

This thesis could not be in its current form if without the academic guidance from my supervisors Prof. Barbro Back and Docent Tomas Eklund. I am indebted to their continuous assistance and encouragement along the way of my doctoral studies.

I would like to thank Akademischer Oberrat Dr. Henning Baars and Prof. Jarmo Partanen for taking time and effort to review my dissertation. I truly appreciate their insightful and constructive comments. I am especially grateful to Akademischer Oberrat Dr. Henning Baars for accepting to act as the opponent at my doctoral defence.

A journey is less enjoyable if without good company. To this end, I would like to express my gratitude to all my collaborators and colleagues. I am particularly grateful to my co-authors Prof. Jüri Vain and Zhiyuan Yao, for sharing my research interests and for cooperating on respective research works. Special thanks also go to Prof. Jüri Vain for hosting my research visit at Tallinn University of Technology, Estonia. I am also grateful of being a member of the Data Mining and Knowledge Management Laboratory at Turku Centre for Computer Science (TUCS). Thank you Minna Tähtinen, Dorina Marghescu, Piia Hirkman, Henri Korvela, Annika Holmbom, Zhiyuan Yao, Peter Sarlin, and Samuel Rönnqvist, for constituting a supportive and inspiring research group, and for the memorable experiences in and outside the ICT-building. Special thanks go to Minna Tähtinen and Piia Hirkman, for helping me overcome the culture and language barriers at many different circumstances; to Annika Holmbom, for generously allow me to use Viscovery SOMine v. 5.0 licenced to her during her maternity leave; to Henri Korvela, for helping me fix the formatting problem in the source data so that I could access them at the beginning of my research and for translating the abstract of my dissertation into Swedish at the end.

I would also like to thank the teaching and administrative staff at the Department of Information Technologies of Åbo Akademi University, and the entire staff at TUCS, for creating a constructive working environment. I particularly appreciate the discussions with Eija Karsten, Karl Rönnholm, Mats Aspnäs, and Tom Fredman at lunch time. I sincerely thank Christel Engblom, Britt-Marie Villstrand, Pia-Maria Kallio, Nina Rytkönen, Tove Österroos, Niklas Grönblom, Joakim Storrank, Tomi Mäntylä, and Irmeli Laine, for various supports I have received in practical matters and for professional services they have provided.

Alas, it is impossible to name all the people who have directly or indirectly influenced this research and have had impacts on the overall experience of my doctoral studies. I am extremely grateful to the case company for providing data for this research (even though the data is not perfect for the research objective). I would also like to thank the Chinese research community in Turku and especially my friends at the Department of Chemical Engineering of Åbo Akademi University, for sharing experiences beyond academic related, and for making me feel that I am connected with my native culture.

I gratefully acknowledge various financial supports I have received during my doctoral studies from different institutions: the Academy of Finland (Grant No. 127656 and 127592), the Finnish Funding Agency for Technology and Innovation (Tekes, Grant No. 40063/08), Fortum Foundation (Scholarship No. 12-099), TUCS, the Department of Information Technologies at Åbo Akademi University, European Social Fund's Doctoral Studies and Internationalization Programme DoRa, Gösta Brander's forskningsfond, Foundation for Economic Education, and Rektor's stipendium.

Save the best for the last. The persons whom I should thank the most are my parents, though they will never expect a thank-you from their children. Their love, support, and understanding have always provided a base camp for me to travel far and explore more. Finally, I would like to thank my husband, Harri Siira, for accompanying me on the 'last mile' of the doctoral journey and for being patient and devoted.

Turku, August 6^{th}, 2014

Table of Contents

List of Figures and Tables

List of Acronyms

ABM	Agent-based Modeling
ACE	Agent-based Computational Economics
AI	Artificial Intelligence
AMI	Advanced Metering Infrastructure
AMR	Automated Meter Reading System
ANN	Artificial Neural Network
BI	Business Intelligence
BMU	best matching unit
BPM	Business Performance Management
CAS	Complex Adaptive Systems
CEER	Council of European Energy Regulators
CEP	Complex Event Processing
CHP	combined heat and power
COLS	Corrected Ordinary Least Square methods
DBMS	Database Management System
DE	Distortion Error
DEA	Data Envelopment Analysis
DG	distributed generation
DM	domain model
DSO	Distribution System Operator
DSR	Design Science Research
DSS	Decision Support Systems
DW	Data Warehouses
EIS	Executive Information Systems
EMA	Energy Market Authority
EU	European Union
HCC	high consumption cluster
IS	Information Systems
ISO	independent system operator
ITO	independent transmission operator
KDD	Knowledge Discovery in Databases
MAMM	multi-agent based meta-model
MAS	Multi-Agent Systems
MDS	Multi-dimensional Scaling
MIS	Management Information Systems
OLAP	Online Analytical Processing

PCA	Principle Component Analysis
QE	Quantization Error
RES	Renewable Energy Sources
RO	research objective
ROR	Rate of Return
SFA	Stochastic Frontier Analysis
SO	research sub-objective
SOM	Self-Organizing Map
TE	Topographic Error
TOU	Time of Use
TSO	Transmission System Operator
UK	United Kingdom
VPP	Virtual Power Players

List of Original Publications

I. **Liu, H.**, Eklund, T., and Back, B. (2012). Smart Metering and Customer Consumption Behavior Profiling: Exploring Potential Business Opportunities for DSOs and Electricity Retailers. In: Jussi Kantola, Waldemar Karwowski (Eds.), Knowledge Service Engineering Handbook, p 179-189, Taylor & Francis, 2012.

II. **Liu**, **H.**, Yao, Z., Eklund, T., and Back, B. (2012). Electricity Consumption Time Series Profiling: A Data Mining Application in Energy Industry. In: Petra Perner (Ed.), Advances in Data Mining: Applications and Theoretical Aspects, LNAI 7377, p 52–66, Springer, 2012.

III. **Liu**, **H.**, Yao, Z., Eklund, T., and Back, B. (2012). From Smart Meter Data to Pricing Intelligence – Visual Data-Mining towards Real-Time BI. In: K. D. Joshi, Youngjin Yoo (Eds.), AMCIS 2012 proceeding, Paper 11, AISeL, 2012.

IV. **Liu**, **H.** and Vain, J. (2013). An Agent-Based Modeling for Price-Responsive Demand Simulation. In: Proceedings of 15th International Conference on Enterprise Information Systems (ICEIS 2013), Paper 67.

V. **Liu, H.**, Eklund, T., Back, B., and Vanharanta, H. (2011). Visual Data Mining: Using Self-Organizing Maps for Electricity Distribution Regulation. In: Proceedings of International Conference on Digital Enterprise and Information Systems (DEIS 2011), July 20-22, London, UK. Springer, CCIS194, p 631-646.

Part I

Research Summary

Chapter 1

Introduction

> The goal of science is to make the wonderful and complex understandable and simple—but not less wonderful.
> --Herbert Simon, *The Sciences of the Artificial*

1.1 What is Business Intelligence

In recent decades, business intelligence (BI) has gained momentum in real-world practice. According to Gartner surveys, 'BI and analytics' has been ranked as the top 1 technology priority by CIOs in 2006-2009, 2012, and 2013, and within the top 5 in 2010-2011. On the one hand, BI has been recognized as a strategic initiative and a key enabler for driving business effectiveness and innovations. On the other hand, the tremendous amount of structured and unstructured data collected in organization's transactional systems, operational systems, and from external data sources and social media has constantly challenged BI practitioners and researchers. The essence of business intelligence deployments is to support better business decision-making. Through handling a large variety of data, BI can help to identify and develop new business opportunities. Making use of new opportunities and implementing an effective strategy will consequently enhance an organization's competitive advantage, and thus, lead to sustainable and profitable growth (Olivia, 2009). In other words, the significance of BI in relation to an enterprise's success is continuously growing in scale, scope, and relevance.

In the meantime, BI has been established as an important research field of Information Systems (IS) within the decision support domain. Evolved from an array of research areas – Decision Support Systems (DSS), Data Warehouses (DW), Executive Information Systems (EIS), and Online Analytical Processing (OLAP), *Business Intelligence* (BI) as it is understood today is an umbrella term to describe a set of methodologies, processes, architectures, technologies, and applications that transform raw data into meaningful and useful information so as to provide actionable insights for business decision makers (Wixom and Watson, 2010).

In particular, business intelligence technologies provide historical, current, and predictive views of business operations. Common functions of business intelligence technologies include querying, reporting, online analytical processing (OLAP), data mining, text mining, process mining, complex event processing (CEP), business performance management (BPM), benchmarking, business analytics, and so on.

It is important to note that in this study business analytics is defined as a subset of BI, instead of as an extension of BI or a separate concept. *Business analytics* has been defined as "a process of transforming data into actions through analysis and insights in the context of organizational decision making and problem solving" (Liberatore and Luo, 2010). In a similar account, Davenport and Harris (2007) pointed out that "analytics are the extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions. …… Analytics are a subset of what has come to be called business intelligence—a set of technologies and processes that use data to understand and analyze business performance."

Basically, business analytics can be categorized into three types (Evans, 2013):

- Descriptive analytics: is to gain insight from historical data through the use of scorecards, clustering, etc., by answering questions such as what has happened, how often, how many, and why it has happened.
- Predictive analytics: is to answer questions such as what will happen next, when it will happen, what is the likelihood of this trend reoccurring, by using predictive modeling, statistical and machine learning techniques.
- Prescriptive analytics: is to recommend decision options and to show the impact of each decision option using optimization, simulation, and so on.

Today's growing competitive pressure in business has led to increased needs for real-time analytics, i.e., so called real-time BI or operational BI (Chaudhuri et al., 2011; Watson, 2009). The value of real-time BI rests in its capability to reduce the three types of latency: *data latency* (the time between business events and when the operational data is captured), *analysis latency* (the time to analyze the data and when the findings are available for use), and *decision latency* (the time to act upon the data) (Hackathorn, 2004). As such, real-time BI can enhance the agility of an organization to significantly increase the responsiveness to varying customer needs and ever-changing market situations.

This is especially true with respect to the electricity production, transmission, distribution, and retail business. Electricity is an increasingly critical infrastructure for post-industrial, information-based economies and societies. Electricity differs from other energy products (e.g., gas and oil) in that it is nearly impossible to be stored economically with current technology, whereas its demand-supply needs to be constantly in balance. The nature of electricity supply determines that the electric systems are among one of the most complex man-made physical systems.

1.2 Why the Electricity Retail Market

Today's electricity sector is subject to complex changes, innovation opportunities, as well as technical and regulatory constraints, ranging from transition to a low carbon society, renewable energy sources (RES) development, market design, to new technologies (e.g., smart metering, smart grids, electric vehicles, etc.), and new independent power producers (e.g., commercial buildings or households with rooftop solar panel installments, a.k.a Distributed Generation). Among them, the ongoing deployment of Advanced Metering Infrastructure (AMI) has profound impacts on the electricity retail market.

Traditionally, electricity consumption data is manually read by the utility personnel at the customer's site on a monthly, quarterly, or annual basis. Both the consumer and the utility can only know the actual amount of consumption one and an half month after the consumption activity has been carried out, in best case. Therefore, the utilities have to estimate their customers' electricity demand based on previous years' consumption patterns. The estimated demand can never be an optimal solution, given that the nature of the electricity supply business requires that electricity producing and consuming activities need to be matched simultaneously at all time. In order to secure the electricity supply (meaning that there is constant power supply whenever an electric device is switched on), and the quality of supply (meaning that the voltage and the frequency are within the variation limits), the power sector has experienced considerable market inefficiency as well as great challenges in production planning, production portfolio optimization, and transmission/distribution network planning and investment. For instance, the investment in reserve capacity is to maintain resource adequacy and system stability when critical peak demand occurs, which usually happens a few times a year and for several hours each time. During the rest of the year, this investment becomes idle and resource-wasteful. The cost of running and maintaining the peak generation and transmission capacity will eventually be transferred to the electricity price, and paid by the consumers.

On the demand side, utilities offer a fixed retail rate that reflects the average cost to serve customers in a broad class over a year or season (e.g., residential customers). This has led to pricing inefficiency and the disconnection of the electricity wholesale and retail markets. Charging the one-size-fits-all class-wide price to customers whose electricity usage patterns can vary widely results in the existing cross-subsidy of high-cost customers by low-cost customers. At the same time, as most consumers are charged according to the fixed retail rate that gives consumers inaccurate information about the actual cost of generating and delivering power, consumers' demand for electricity remains independent of conditions in the wholesale markets. Therefore, they have no incentive to adjust electricity usage when wholesale prices rise above or fall below the fixed retail rate.

With the implementation of remotely-readable, two-way communication, and automated meter reading systems (i.e., smart meter, AMI), the electricity consumption data can be captured in 15- to 60-minute time intervals. Not only can consumers be informed about their electricity consumption and the varying wholesale price in a timely manner, the utilities can also get access to customers' demand patterns in near real-time. On the one hand, this will support efficient pricing, which reflects time-varying wholesale prices, diverse consumer risk preferences, and relevant cost to suppliers. On the other hand, it will facilitate greater involvement of the demand side. It is projected that the deployment of AMI opens up opportunities for an active demand response electricity retail market, which will enable close interaction between electricity wholesale and retail markets, and thereby improve overall market efficiency[1].

1.3 Motivation and Research Objectives

From the view point of BI research, the AMI is enabling real-time or near real-time analytics in the electricity retail business. Nevertheless, the roll-out of smart meters is merely one step further on the way to tackle the *data latency* issue. How can massive volumes of smart meter data be turned into actionable insights? How can electricity consumption behavior profiling (i.e., instant and in-depth customer analysis and/or advanced segmentation according to real-time energy consumption patterns) provide supporting information in pricing decision-making? How can customers' demand be steered in response to the electricity wholesale market price or when the electric systems stability is at

---

[1] Efficiency means that resources are not being wasted. Economists identify two types of wastage: (i) physical wastage, in the sense that some valued units of resource remain unused; and (ii) wastage of value, in the sense that some units of resource are not being used by those who value them most. The concept of 'market efficiency' addresses both (i) and (ii). (Tesfatsion, 2009)

risk (i.e., demand response)? These questions are in relation to the energy consumption data utilization challenges the electricity sector is facing. Based on such understanding, this research aims at exploring some aspects of BI to address the *analysis latency* issue, with a particular focus on retail pricing.

In a competitive electricity market, electricity retail rates basically consist of three major components – energy prices, network tariffs, and taxes/levies. Energy prices reflect the electricity prices in the wholesale market and service costs of the electricity suppliers (retailers). Network tariffs reflect the operational efficiency, the investment efficiency, and the operational environment of the distribution network operators (DSOs). Taxes and levies are based on governmental policies. In other words, from the view point of the electricity industry, retail pricing concerns: (1) the electricity supplier's pricing practice; and (2) the DSO's efficient performance by means of benchmarking. Therefore, this research covers both. *Innovative retail pricing*, enabled by the AMI deployment, is from the perspectives of electricity suppliers. E*fficiency benchmarking*, as a relevant concern with respect to the retail price and the electricity retail market, is included and is from the perspective of electricity distribution and regulation. In terms of innovative retail pricing, the research focus is on consumption behavior profiling, real-time BI architecture, and price-responsive demand modeling and simulation. In terms of efficiency benchmarking, the research focus is on efficient performance visualization.

Hence, the research objectives are:

- to use visual data mining techniques to perform customer consumption behavior profiling (**RO1**);
- to develop a real-time BI framework and a price-responsive demand modeling method for dynamic pricing and demand response applications (**RO2**);
- to visualize the efficient performance of DSOs (**RO3**).

This study is from the perspective of the European electricity industry, particularly focusing on the Nordic countries (e.g., Finland and Estonia). One reason is that the European Union (EU) countries have adopted an unbundled electricity market structure (which will be explained in Chapter 3). This gives rise to the potential impact of this research on a Pan-European scale. The second reason is that when this research was initiated, Finland is one of the three European countries (i.e., Italy, Sweden, and Finland) who have made the decisions to roll out smart meters as of 2011. Also getting empirical data for this research is a practical reason.

The specified research sub-objectives include:

SO1. To explore the potential value of the Self-Organizing Map (SOM) method in electricity consumption profiling;

SO2. To investigate the potential of Agent-based Modeling (ABM) in price-responsive demand simulation;

SO3. To demonstrate how a real-time BI approach can contribute to obtaining and maintaining an innovative and demand response-oriented electricity retail market in the long run;

SO4. To examine the utility of using the Self-Organizing Map (SOM) for efficiency benchmarking.

This research predefines visual data mining as one of the solution objectives. The goal of data mining is to extract hidden knowledge and patterns in large amounts of data so as to provide support for decision making, while visualization is to represent the data and the data mining results in a understandable and intuitive way so as to aid visual exploration of the useful patterns in the data (Ferreira de Oliveira and Levkowitz, 2003; Keim, 2001). Hence, the approach of combining data mining and visualization techniques, i.e., visual data mining, has been contemplated for the current research tasks (**RO1** and **RO3**).

In particular, the motivation for selecting the SOM method in the specific application domain (**SO1** and **SO4**) lies in that the SOM is an artificial neural network (ANN)-based, unsupervised, data mining method, and it enables data and dimension reduction simultaneously. Compared to several popular algorithms such as *k*-means, Fuzzy c-means, and Vector Quantization for clustering, and Principal Component Analysis (PCA), Multi-dimensional Scaling (MDS), and Sammon's mapping for visualization, the SOM is suitable for both clustering and visualization tasks. A further discussion of the data mining concept and the SOM method is given in Section 2.2.

The motivation for proposing ABM for price-responsive demand simulation (**SO2**) is due to that the electricity retail market is considered as a complex economic system, and ABM is a promising modeling approach in studying complex adaptive systems (CAS). ABM allows to re-create or predict the emergence of complex phenomena via simulating the actions and interactions of multiple autonomous agents in a computational setting, in order to assess the impacts on the system before any real-world implementations. This property satisfies the requirements in designing dynamic pricing and demand response applications. A detailed discussion is given in Section 2.3.

1.4 Research Design

The field of this research has an inter-disciplinary nature, i.e., it draws upon research in computer science, economics, engineering, and information systems. The current research is initiated by recognizing the challenges and opportunities the electricity retail market is facing with the implementation of smart meters, and by the understanding of the strategic importance of BI in relation to the success of today's organizations. In the first step, a set of SOM-based visual data-mining models were investigated in order to profile the customers' electricity consumption patterns in the case company. Accordingly, the real-time BI enterprise architecture is proposed based on empirical findings. Next, given that the enterprise-wide real-time BI infrastructure is in place, a multi-agent-based modeling is demonstrated. To a certain extent, this is served to justify the proposed BI architecture. In addition, a data-mining application which addresses efficiency benchmarking in electricity distribution regulation is included as a related topic in the electricity retail market.

Hence, the building blocks of this research consist of:

(i) visual data mining-based descriptive analytics, namely electricity consumption profiling, for pricing decision-making support;
(ii) real-time BI enterprise architecture for enhancing management's capacity on real-time decision-making;
(iii) prescriptive analytics through agent-based modeling for price-responsive demand simulation;
(iv) a visual data-mining application for electricity distribution benchmarking.

1.5 Structure of this Dissertation

This thesis is the consolidation of a collection of studies. It constitutes two parts – Part I, is the research summary, which highlights the essentials of the research topics under study and ties the individual research into one piece; Part II, consists of reprints of original double-blind peer-reviewed publications, which provides supporting information regarding each individual study.

The research summary (Part I) is organized as follows. **Chapter 1** presents the research background, motivation, objectives, and the research design. **Chapter 2** is devoted to research methodology and methods, including the

Design Science Research (DSR) paradigm, the concept of Knowledge Discovery in Databases (KDD) and data mining, the Self-Organizing Map (SOM), as well as the theory of Complex Adaptive Systems (CAS) and Agent-based Modeling (ABM). **Chapter 3** is focused on the state-of-the-art in the electricity industry, where the deregulation of the electricity markets, the concepts and current practices around smart metering, smart grids, and demand response, will be introduced. Meanwhile, the relevance to the IS research field, i.e., related studies regarding data mining in the energy industry and agent-based modeling in electricity market analysis, is reviewed. From Chapter 4 throughout Chapter 7, the four aspects of BI under study are elaborated. **Chapter 4** presents visual data-mining based descriptive analytics, namely electricity consumption profiling. **Chapter 5** discusses enterprise-wide real-time BI architecture and its implications for an electricity utility. **Chapter 6** demonstrates prescriptive analytics based on multi-agent simulation, which also serves as a proof-of-concept. **Chapter 7** touches upon electricity distribution benchmarking through SOM-based clustering. Finally, **Chapter 8** concludes the thesis with research contributions, limitations, and future research directions.

The thesis is based on five co-authored publications, which are included in Part II. The publications and the author's contribution in each publication are as follows:

Publication I: Liu, H., Eklund, T., and Back, B. (2012). Smart Metering and Customer Consumption Behavior Profiling: Exploring Potential Business Opportunities for DSOs and Electricity Retailers. In: J. Kantola, and W. Karwowski (Eds.), Knowledge Service Engineering Handbook, p 179-189, Taylor & Francis, 2012.

<u>The author's contribution</u>: As the main author, I initiated the research idea and carried out the modeling and analysis tasks. The co-authors involved in results analysis and refining the paper.

Publication II: Liu, H., Yao, Z., Eklund, T., and Back, B. (2012). Electricity Consumption Time Series Profiling: A Data Mining Application in Energy Industry. In: Petra Perner (Ed.), Advances in Data Mining: Applications and Theoretical Aspects, LNAI 7377, p 52–66, Springer, 2012.

<u>The author's contribution</u>: I initiated the research idea and cooperated with Yao on modeling. Yao was responsible for the data pre-processing and modeling, while I was responsible for the analysis. Yao and I were responsible for writing most of the manuscript, while other co-authors involved in results evaluation and refining the paper.

Publication III: Liu, H., Yao, Z., Eklund, T., and Back, B. (2012). From Smart Meter Data to Pricing Intelligence – Visual Data-Mining towards Real-Time BI. In: K. D. Joshi, Youngjin Yoo (Eds.), AMCIS 2012 proceeding, Paper 11, AISeL, 2012.

The author's contribution: As the first author, I initiated the research idea and developed the real-time BI framework based on previous study. Other co-authors contributed to refining the BI framework and the manuscript.

Publication IV: Liu, H. and Vain, J. (2013). An Agent-Based Modeling for Price-Responsive Demand Simulation. In: Proceedings of 15th International Conference on Enterprise Information Systems (ICEIS 2013), Paper 67.

The author's contribution: I initiated the research idea and collaborated with Vain on modeling. Vain implemented the simulation, and I was responsible for the analysis. Even though I am the first author, Vain has contributed an equal share.

Publication V: Liu, H., Eklund, T., Back, B., and Vanharanta, H. (2011). Visual Data Mining: Using Self-Organizing Maps for Electricity Distribution Regulation. In: Proceedings of International Conference on Digital Enterprise and Information Systems (DEIS 2011), July 20-22, London, UK. Springer, CCIS194, p 631-646.

The author's contribution: Eklund and I initiated the research idea, and I carried out the modeling and analysis tasks. I am the main author. The co-authors were involved in results analysis and refining the paper.

Figure 1 shows an overview of the thesis.

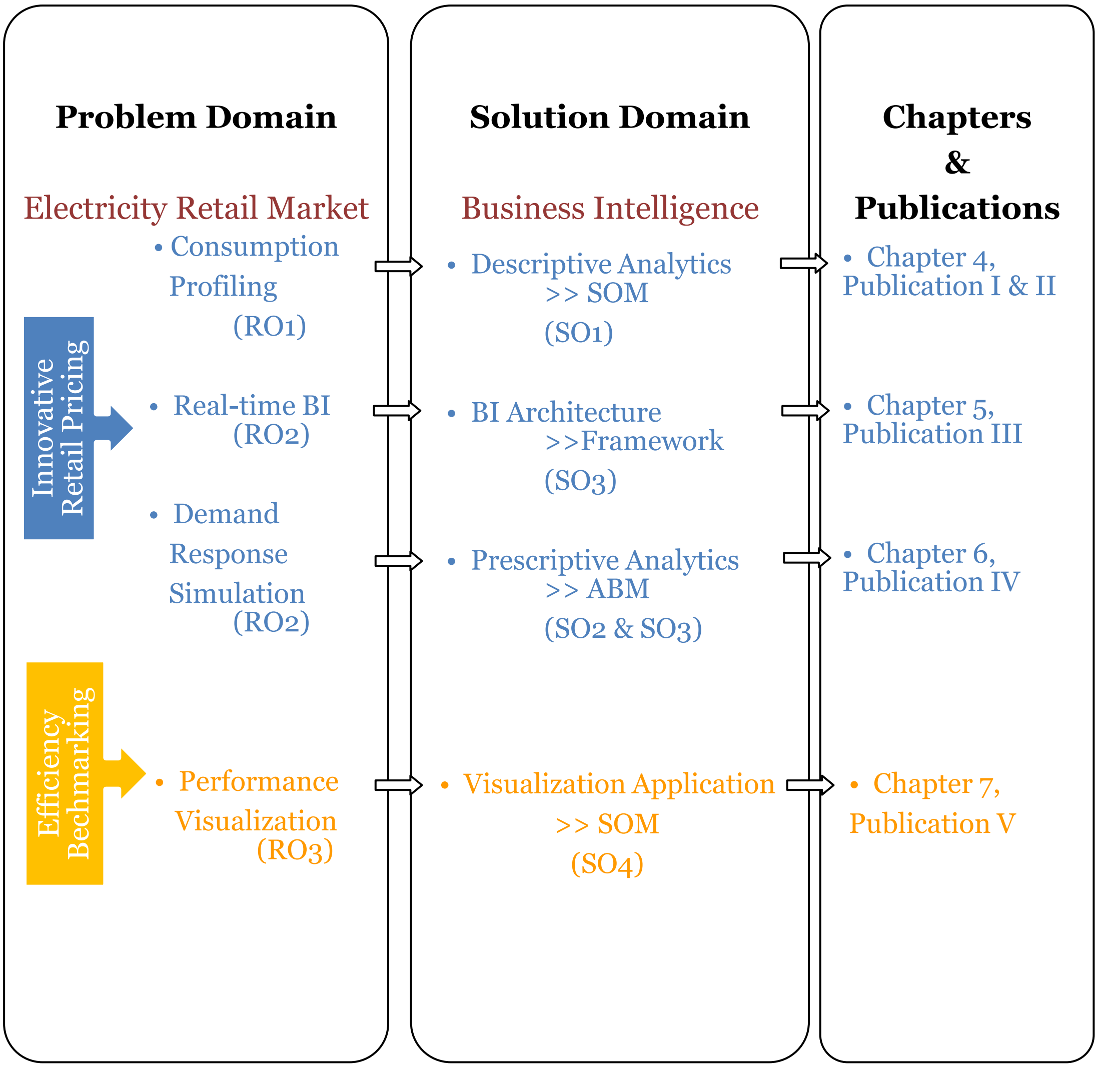


Figure 1. An Overview of the Thesis

Chapter 2

Research Methodology and Methods

Research methodology is the scientific foundation according to which the research has been built. This chapter is devoted to the research methodology and the methods employed in this research. In the first section, the design science research (DSR) paradigm in the IS field and the rationale for adopting the DSR in this research will be introduced. This is to anchor the current research in the scientific undertaking. In the second section, the principles of Knowledge Discovery or Knowledge Discovery in Databases (KDD) and data mining in general, and the Self-Organizing Map (SOM) as a visual data mining method in particular, will be presented. In the third section, Agent-based Modeling (ABM) as one type of computational intelligence in studying complex adaptive systems (CAS) will be discussed.

2.1 IS Research Paradigm: Design Science Research

Information Systems (IS) research is an applied discipline. It draws on a number of other fields for its motivations, techniques, and applications, such as computer science, system science, management science, social science, and economics. There is a broad consensus that the core of the discipline emphasizes the development and use of IT artifacts in organizations (Benbasat and Zmud 2003; Iivari 2003, 2007; Orlikowski and Iacono 2001). This gives rise to Design Science Research (DSR) as an important paradigm of IS research.

DSR, as conceptualized by Simon (1996) in '*The Sciences of the Artificial*', takes a construction-oriented view of IS research. As Hevner et al. (2004) articulated, DSR is centered around designing and building innovative IT artifacts, as well as developing design theory relevant for practice. As such, DSR aims to address the dual missions of IS research: make theoretical contributions and assist in solving current and anticipated problems of practitioners (Benbasat and Zmud 1999; Iivari 2003; Rosemann and Vessey 2008). In other words, DSR is a solution-oriented and prescription-driven paradigm (van Aken, 2004), which aims at contributing to the knowledge body of IS practice and research via constructing IT artifacts combining novelty and utility (March and Storey, 2008). The design artifact must serve identified important business opportunities (Iivari, 2007) and/or provide solutions to challenging management problems (Gregor and Jones, 2007). These characteristics motivated the author to adopt the DSR paradigm in this research.

A number of IS scholars (Hevner et al. 2004; March and Smith 1995; Nunamaker et al. 1991; Peffers et al. 2008; Walls et al. 1992) have elaborated on the design science research method in various forms. Essentially, the design science research activities are mainly organized around the course of *build-evaluation-theorize-justify*, and the four phases of research activities are iterative and interactive across the *relevance cycle*, the *rigor cycle*, and the *design cycle*. As in Hevner (2007), the relevance cycle initiates the requirements from the contextual environment of the application domain and returns the research artifacts into environment for field testing. The rigor cycle provides the foundations for rigorous design science research from a vast knowledge base including grounding theories, methods, and domain expertise, and extends the knowledge base with new knowledge generated by the research activities. The design cycle iterates between the activities of building and evaluating the design artifacts until satisfying the requirements. Figure 2 illustrates the three-cycle view of DSR.

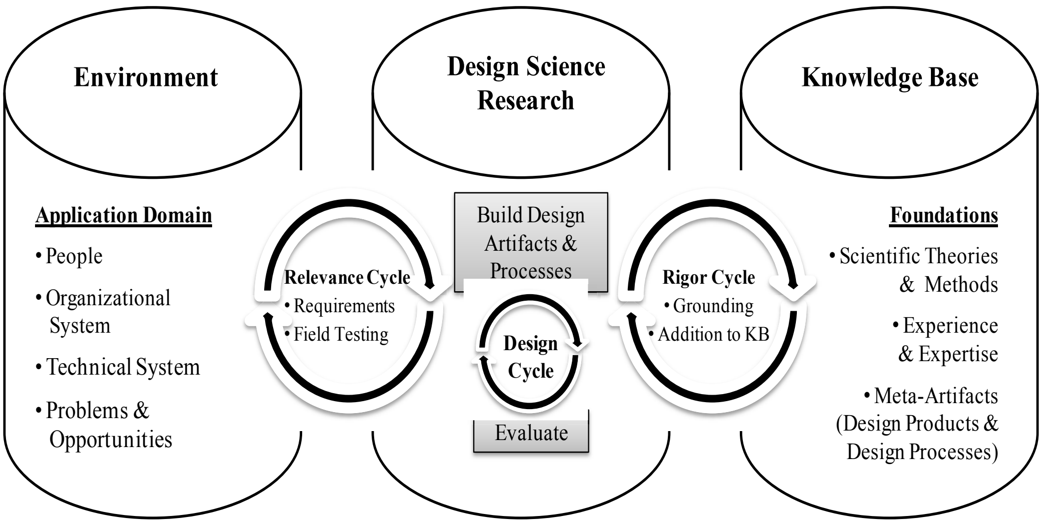


Figure 2. Design Science Research Cycles (adopted from Hevner, 2007)

The contributions of DSR to the knowledge base are dominated by prescriptive knowledge. Descriptive knowledge, which constitutes the bulk of the natural and social sciences, provides the theoretical bases for the design of successful artifacts. Prescriptive knowledge is concerned with 'socio-technical artifacts', i.e., the sciences of the artificial (Gregor and Hevner, 2013).

Following March and Smith (1995), and Gregor (2006), prescriptive knowledge can be categorized into different forms:

- *Constructs*, which are concepts and symbols used to define and understand the problem and solution domains.
- *Models*, which are the representations of possible solutions to the identified problem.
- *Methods*, which are algorithms, techniques, and technological rules for performing goal-driven activities.
- *Instantiations*, which are physical systems, products, and processes where design knowledge can be inferred to some degree.
- *Design theory*, which is an abstract statement that provides formalized prescriptions for designing and developing an artifact.

This research is conducted in parallel with the DSR process described by Peffers et al. (2008), including the following steps:

- Step 1: problem identification and motivation;
- Step 2: define solution objectives;
- Step 3: design and development;
- Step 4: demonstration;
- Step 5: evaluation;
- Step 6: communication.

As mentioned earlier, this research is initiated by recognizing the business opportunities and challenges that are enabled by the implementation of AMI in the problem domain of electricity industry (step 1). In other words, the research process starts from the relevance cycle. Accordingly, a real-time BI framework based on descriptive analytics and prescriptive analytics is proposed (step 2). This is done by reviewing IS literature related to BI, empirical studies in the application domain, and both theoretical and application oriented literature regarding data mining and agent-based modeling techniques (see e.g., Section 3.4 and Section 5.1). The existing knowledge formed the scientific underpinnings to build the data-mining model, the real-time BI architecture, and the agent-based modeling and simulation in the design cycle (step 3). In terms of the overall research design, the agent-based modeling and simulation also serves to demonstrate the real-time BI framework under study (step 4). During the respective processes of data mining and agent-based modeling, technical evaluation (model verification) is performed (step 5). It is important to note that the rigor cycle spans from step 2 to step 5. In principle, the technical evaluation and the demonstration functioned as proof-of-concept. Further proof-of-value and proof-of-use analyses require field testing, which will lead to the return to the relevance cycle through communication (step 6). At the current stage, the communication is in the form of peer-reviewed publications, i.e., mainly to the academic community; meanwhile, certain research results have been presented at electricity industry focused professional conference such as CIRED (International Conference on Electricity Distribution) in 2011.

2.2 KDD, Data Mining, and the SOM

As stated in Chapter 1, visual data mining is predefined as one of the solution objectives (**RO1**), and the Self-Organizing Map (SOM) is selected as the data mining method due to its property of simultaneous visualization and clustering (**SO1** and **SO4**). Here the science of knowledge discovery or knowledge discovery in Databases (KDD) is the theoretical underpinning for deriving these solution proposals. In the following, the core concepts of KDD and data mining, and the SOM method will be presented.

> - to use visual data mining techniques to perform customer consumption behaviour profiling (**RO1**);
>
> **SO1.** To explore the potential value of the Self-Organizing Map (SOM) method in electricity consumption profiling;
> **SO4.** To examine the utility of using the Self-Organizing Map (SOM) for efficiency benchmarking.

2.2.1 Knowledge Discovery through Data Mining

It is often claimed that the terms *knowledge discovery* and *data mining* are interchangeable (Ronald, 2001), as both refer to the discovery of useful patterns in data. Historically, the term *knowledge discovery* is more popular in the artificial intelligence (AI) and machine learning fields, while the term *data mining* is mostly used by statisticians, data analysts, and the database and management information systems (MIS) communities. However, since the inception of the concept of *knowledge discovery in databases* (KDD) in 1989, the distinction between knowledge discovery or KDD and data mining is identified and acknowledged within scientific communities. Knowledge discovery, or KDD, as we know it nowadays, refers to the *overall process* of discovering 'valid, novel, potentially useful, and ultimately understandable patterns in data', whereas data mining refers to a *particular analysis step* in this process (Fayyad et al., 1996).

There are various versions of the KDD process, ranging from the nine-step model by Fayyad et al. (1996), the eight-step model by Anand and Buchner

(1998), the six-step model by Cios et al. (2000), the five-step model by Cabena et al. (1998;), to the six-step industry model CRISP-DM (Shearer, 2000). Nevertheless, the core of the KDD process can be summarized into six major steps, including: (1) understanding the application domain; (2) understanding the data; (3) data preprocessing and choosing data mining method; (4) data mining/modeling; (5) evaluation; and (6) consolidating and deploying the discovered knowledge. Even though the steps are executed in sequence, the overall KDD process is highly iterative, and includes many feedback loops and repetitions.

In particular, according to ACM SIGKDD[2], *data mining* is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. The actual data mining task is the automatic or semi-automatic analysis of large quantities of data in large relational databases to extract implicit or previously unknown, yet interesting patterns such as groups of data records (cluster analysis or clustering), unusual records (anomaly detection), dependencies (association rule mining), categories for identifying new observations (classification), and so on.

In addition, visualization is very often integrated in the KDD process. In the context of KDD, visualization can be classified into *visualization of the data* and *visualization of the data mining results* (Ferreira de Oliveira and Levkowitz, 2003). Statistical graphics such as histograms, scatter plots, bar/pie/line charts, etc. are normally applied in step (2) for understanding the data, i.e., visualization of the data. In order to facilitate the visual exploration of discovered patterns intuitively according to human cognition, data mining results need to be displayed in an easily understandable and effective manner, i.e., visualization of the data mining results. Usually, the choice is to combine data mining with visualization techniques, or to select the method which is suitable for data mining and visualization tasks simultaneously, thus, to conduct *visual data mining*.

Rooted in the advances of KDD, given the context of massive volumes of timely measured smart meter data, accordingly, the research interest emerged – discovering implicit electricity consumption patterns by means of visual data mining, i.e., knowledge discovery through data mining.

[2] ACM SIGKDD is the Association for Computing Machinery's (ACM) Special Interest Group (SIG) on Knowledge Discovery and Data Mining (KDD).

It is necessary to note that there are a variety of methods/techniques involved regarding each step of the KDD process. Thoroughly discussing them, however, is not the focus of this study. The following part is concentrated on the data mining method employed in this research, i.e., the Self-Organizing Map (SOM).

2.2.2 The SOM Method

In this subsection, first, the key properties of the SOM and the motivation of choosing the SOM for the research tasks will be introduced. Then, the working mechanisms in training, evaluating, and visualizing the SOM, will be presented.

Key Properties

The SOM is a data mining method based upon artificial neural networks (ANNs). ANNs are designed to mimic the basic learning and association patterns of the human nervous system, and consist of a number of neurons (simple processors) connected by weighted connections. ANNs learn by adjusting the weight of each connection, increasing or decreasing the importance of the input (information) being transferred, until a desired output is achieved. Essentially, they are nonlinear, multivariate regression techniques, better able to handle erroneous and noisy data than parametric statistical tools (Bishop, 1995). Like other ANNs, the SOM is acknowledged for its robustness in handling nonlinear and multivariate data, especially with respect to dimension reduction and large datasets. The SOM is considered particularly suitable for clustering and visualization tasks (Han and Kamber, 2000).

The SOM is a two-layer ANN (i.e., an input and an output layer) that applies the *unsupervised learning* paradigm, meaning that the SOM does not require target output values for training (Kohonen, 2001). In other words, the SOM is a *clustering* algorithm, which divides data records into groups based on similarities and differences in a collection of unlabelled data. Since unsupervised learning is viewed as spontaneously finding patterns and structure in the input data space, the SOM method can be described as descriptive. Yet, the SOM could be extended to perform predictive tasks, as unsupervised learning can be a precursor to supervised learning (Abu-Mostafa et al., 2012)

From the perspective of dimension reduction, the SOM belongs to a class of topology-preserving based projection methods. The essence of the SOM is to map high dimensional data onto a spatial map (usually in the form of a two-dimensional grid made up of hexagonal lattices). The SOM uses the *competitive learning* algorithm, meaning that the reference vectors on the output layer compete with each other to be the best matching unit (BMU, i.e., the winner) whose connection weights to the input vectors are the closest in terms of the *Euclidian distance*. At the same time, the SOM algorithm allows the reference vectors in the neighborhood of the BMU to adjust their weights accordingly. Theoretically, all the nodes on the output layer are a projection of the input data items. As such, the intrinsic relationships (e.g., similarities) of input data in the multivariate space are reflected on a two-dimensional topological map, that is, *visual clustering* is performed (Haykin, 1999; Kohonen, 2001).

Based on the key properties, the reasons for choosing the SOM in electricity consumption profiling and for distribution efficiency benchmarking are explained as follows:

- simultaneous data reduction (i.e., clustering) and dimension reduction (i.e., projection) capabilities;
- robustness in handling data with missing value;
- computational efficiency;
- well-documented and received by research communities for broad application domains.

Firstly, the SOM's capability of performing clustering tasks and simultaneously providing a multivariate visual display of the clustering results makes it a suitable method for visual data mining. In addition, compared to similar clustering techniques (such as *k*-means), the SOM does not require predefining a desirable number of clusters, which means that little *a priori* knowledge of the data is required. In terms of the effectiveness of preserving the structure of the original data, the SOM's neighborhood preservation capability is advantageous, in comparison to other projection methods such as Multi-dimensional Scaling (MDS), Principle Component Analysis (PCA), and Sammon's mapping (Venna and Kaski, 2001; Marghescu, 2006).

Secondly, as mentioned above, the SOM as one type of ANN is robust in handling data with missing values (Kohonen, 2001), a phenomenon that happens very often with respect to time-series data. Unlike certain algorithms such as PCA or *k*-means that disregard the data records with missing values completely, the SOM can still utilize the available information in those data records in training.

Thirdly, given large volume of data, computational efficiency also needs to be taken into account when selecting the appropriate data mining method. The SOM has relatively sound computational efficiency since it conducts clustering and projection tasks simultaneously. In order to improve the computational efficiency further, the batch training of the SOM is used in this research, which will be discussed in the following of this subsection.

Last but not the least, the SOM has been extensively studied in applications in finance, medicine, and engineering (Deboeck and Kohonen, 1998; Oja et al., 2002). For instance, the SOM has been used for financial performance analysis and benchmarking (Serrano-Cinca, 1996; Back et al., 1998; Eklund et al., 2008), customer portfolio analysis and segmentation (Denny et al., 2010; Holmbom et al., 2011; Yao et al., 2010), and currency crises analysis (Sarlin, 2011). With respect to the energy sector, the SOM has been used for electric power system stability assessment and online-provision control (Nababhushana et al., 1998), and for visualization of voltage stability in a large power plant (Rehtanz, 1999). Lendasse et al. (2002) and Riqueline et al. (2000) examine the possibility of forecasting electricity consumption/short-term load with the SOM, and Kim et al. (2000) report electric load forecasting and classification using the SOM. This shows that the usefulness of the SOM in broad application domains has been well received by research communities.

Motivated by these key features of the SOM, this study seeks to investigate the potential value/utility of the SOM in the application domain of electricity consumption behavior profiling and electricity distribution regulation benchmarking. The SOM-based data-mining approach is intended to function as an option of descriptive BI solution in the context of the electricity retail market (see Chapter 4).

Hereafter, the working mechanisms of the SOM, which are relevant to the current research, will be briefly introduced.

Training the SOM

Firstly, before the SOM algorithm is initiated, the topological structure is defined. In practice, the form of two-dimensional grid made up of *hexagonal lattices* is recommended due to the advantages in terms of interpretability, visual appealing, and six neighbors of each node being at the same distance as opposite to the eight neighbors in rectangular lattice (Kohonen, 2001; Vesanto et al., 1999). Secondly, the *initialization* of the reference vectors (e.g., random, sample, or linear initialization), and the *neighborhood function* (e.g., Gaussian), are specified. In practice these specifications are very often bounded by the software package chosen. Thirdly, the map training involves iterating a series of parameter combinations, including the number of *nodes* (i.e., the size of the map), the *map shape* (i.e., the ratio of X and Y dimensions), the *radius* of the neighborhood, the *learning rate* in the case of the sequential training, etc.

The batch training algorithm is commonly adopted, instead of the sequential training algorithm. In sequential training, the weights of the network are updated each time an input is mapped. Whereas, in batch training, the weights of the network are only updated after all the input data are mapped. Therefore, the batch training is much more computationally efficient than the sequential training since the map is updated less often. And it makes the maps trained with different parameter combinations comparable since the map training is independent of the order of the data given the same initialization. It is necessary to note that in batch training the learning rate does not need to be specified as in the case of the sequential training, yet, the neighborhood function and the radius of the neighborhood have to be defined.

The SOM algorithm can be expressed mathematically as follows:

Step 1: Mapping the input vectors to corresponding BMUs

$$c(x_i) = \arg\min_i\{\|x_i - m_k(t)\|\} \tag{2.1}$$

Step 2: Batch update the weights of the reference vectors in the neighborhood

$$m_k(t+1) = \frac{\sum_{i=1}^{N} h_{c(x_i)k}(t)x_i}{\sum_{i=1}^{N} h_{c(x_i)k}(t)} \tag{2.2}$$

where:

$c(x_i)$ denotes the BMU of input vector x_i (i = 1, 2, ... N);

$m_k(t)$ denotes the reference vector of each input vector at time t (k = 1, 2, ... M);

‖.‖ denotes the Euclidean distance;

$h_{c(x_i)k}$ denotes the Gaussian neighborhood function.

$$h_{c(x_i)k} = exp\left(-\frac{\|r_c - r_i\|^2}{2\sigma^2(t)}\right) \tag{2.3}$$

where:

r_c denotes the coordinate of the BMU;

r_i denotes the coordinate of the reference vector m_i;

$\sigma(t)$ denotes the radius of the neighborhood, t is a discrete time coordinate.

In addition, a two-level clustering, i.e., the SOM-Ward clustering, is used in this research. First, the data is projected/grouped onto a two-dimensional display using the SOM. Then the resulting SOM is clustered using Ward's clustering. Belonging to a class of hierarchical clustering methods, Ward's

(1963) clustering method attempts to minimize the total within-cluster variance, meaning that the pair of clusters with minimum between-cluster distance is merged. The pairwise distance metric of Ward's method is the squared Euclidean distance of cluster centroids. As such, the final clustering results are optimized with high intra-cluster similarity and low inter-cluster similarity.

Evaluating the SOM

Like in any data mining process, training and evaluating the trained/mined results are performed iteratively and interactively. The most commonly used goodness measures for evaluating the SOM are: *Average Quantization Error*, *Distortion Error*, and *Topographic Error*.

Average Quantization Error (QE) measures the resolution of the map, which computes the average distance between each data vector (x_i) and its BMU ($c(x_i)$). QE can be computed as in (2.4). It is worth to note that for any given dataset, the QE can be reduced by simply increasing the number of map nodes. However, doing so usually leads to distortion of the map's topology (Pölzlbauer, 2004).

$$\varepsilon_{qe} = \frac{1}{N} \sum_{i=1}^{N} \min\{\|x_i - c(x_i)\|\} \tag{2.4}$$

Similar to QE, *Distortion Error (DE)* is used to measure the quality of vector quantization. Since the distortion measure is taking the neighborhood into account, DE can be used to select the best fitting map from several maps trained with the same dataset (Pölzlbauer, 2004). A mathematical expression of DE is as follows:

$$\varepsilon_{de} = \frac{1}{N}\frac{1}{M} \sum_{i=1}^{N} \sum_{k=1}^{M} h_{c(x_i)k} \|x_i - c(x_i)\|^2 \tag{2.5}$$

Topographic Error (TE) measures the continuity of the mapping, i.e., the topology preservation, which computes the proportion of all data vectors for which the first and the second BMUs are not adjacent units (Kiviluoto, 1996). A mathematical expression of TE is as follows:

$$\varepsilon_{te} = \frac{1}{N} \sum_{i=1}^{N} u(x_i), \tag{2.6}$$

where $u(x_i) = \begin{cases} 1, & \text{if 1st and 2nd BMUs are not adjacent} \\ 0, & \text{otherwise} \end{cases}$

Visualizing the SOM

The final clustering results of the SOM can be presented in various forms for visual exploration, such as Unified distance matrix (U-matrix) (Ultsch and Siemon, 1990), frequency plots, error plots, component planes (also known as feature maps). Component planes are commonly used when a lot of information (i.e., variables) has to be visualized at once (Vesanto et al., 2000). A component plane represents each input variable with the same predefined SOM grid structure and the same map size. The order of the component planes follows the order of the variables as they were presented in the datasets. The node in the same location across the component planes represents the same input data. As such, a set of component planes provides a multivariate view of a particular input. In addition, the variables that are used for training the map are usually displayed in color, with 'warm' colors (e.g., red) representing high values while 'cold' colors (e.g., blue) represent low values. Therefore, it is easy to visually interpret the characteristics of each cluster from the component planes. For instance, Figure 3 displays the component planes of load profile at four different time periods within 24 hours (i.e., 6:00-9:00, 9:00-16:00, 16:00-22:00, 22:00-6:00). Here each time period represents an individual variable. It shows that high values at the upper left corner of the component planes belong to Cluster I, whereas low values on the right side of the component planes belong to Cluster IV. Moreover, the peak demand of Cluster I during the time period 6:00-9:00 and 22:00-6:00 are relatively higher, compared to those during the time period 9:00-22:00.

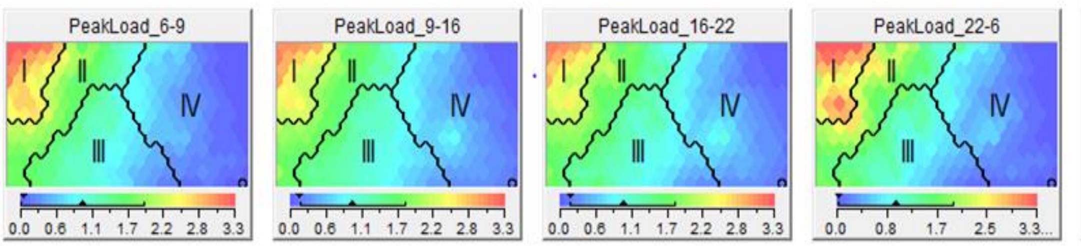


Figure 3. An Example of Component Planes (based on Publication II)

2.3 ABM in Complex System Simulation

As stated in Chapter 1, Agent-based Modeling (ABM) is another solution objective which is under study. Hence, this section will first introduce the theoretical foundation – Complex Adaptive Systems (CAS), and then present the agent-based modeling (ABM) methodology.

- ➢ to develop a real-time BI framework and a price-responsive demand modeling method for dynamic pricing and demand response applications (**RO2**);

SO2. To investigate the potential of Agent-based Modeling (ABM) in price-responsive demand simulation;

2.3.1 Complex Adaptive Systems (CAS)

The concept of Complex Adaptive Systems (CAS) has evolved from studies of complexity and system science since the mid-1990s. Coined by the scholars at the Santa Fe Institute, *complex adaptive systems* are defined as systems that have a large numbers of components, often called agents, which adapt or learn as they interact with each other and with the environment surrounding them (Holland, 2006). Typical examples of complex adaptive systems include economies, stock markets, ant colonies, ecosystems, the brain and the immune system, a supply chain, and any human social group such as political parties, communities, geopolitical organizations, and so on. The study of CAS is highly interdisciplinary – that is, it is closely associated with the field of Artificial Intelligence (AI) and Multi-Agent Systems (MAS) in computer science, but it has a primary focus on social systems. The study of CAS has also drawn on the insights from general systems theory, chaos theory, catastrophe theory, game theory, and the scientific discourses from evolutionary algorithms and cellular automata, etc.

The key principles of CAS include *emergence* and *self-organization*. It is important to note that emergence and self-organization appear in a variety of physical, chemical, biological, social, and cognitive systems. Thus, the conceptualization of emergence and self-organization may slightly vary in different disciplines. The discussion hereafter is from the perspective of social systems.

Emergence

Emergence refers to collective phenomena that are collaboratively created by individuals, yet are not reducible to individual action (Sawyer, 2005). In other words, the property of the whole (e.g., stock market) is greater than that of the simple sum of its constituent components (e.g., individual traders/investors). The macroscopic patterns, regularities, and properties in a complex system generated by a number of decentralized, or distributed, individual entities operating in an environment are called *emergent behavior* (e.g., stock market prices). The emergent behavior is not properties of any single such entity, nor can it easily be predicted or deduced from the behavior of the micro-level entities. CAS regularly exhibits emergent behavior, such as standing ovations (Schelling, 1978), trade networks, and traffic jams.

Emergent behavior in a CAS is difficult to predict because agents' interactions are not independent and the feedback can be positive and/or negative which will alter the dynamics of the system. When the agent's number is in the order of multitude, in a system with negative feedback, individual variations will cancel one another out, and therefore, changes get absorbed quickly and the system stays stable. On the contrary, with positive feedback, changes become amplified or reinforced so that the system experiences instability (Miller and Page, 2007). When a system contains both types of feedback, which is very common in many CAS, novel global phenomena may possibly emerge as the result of local entities' interactions. In addition, in many social systems, agents have the capability to adjust their behavior based on past experience (e.g., positive or negative outcomes) or in anticipation of future events. Agents' adaptation will alter the dynamics of the system, which in turn will result in unexpected emergence.

Self-organization

On the other hand, CAS also exhibits the property of self-organization. Self-organization refers to coherent processes in which the micro-level components interact in a well-regulated yet spontaneous fashion so that macroscopic patterns, structures, and phenomena are assembled, aggregated, and emerged. Adam Smith's 'invisible hand' is a typical denotation of self-organization of collections of self-interested agents leading to well-structured free markets. Yet, 'what are the self-organizing capabilities of decentralized market economies?' still remains unresolved for agent-based computational economists (ACE) to look into (Tesfatsion and Judd, 2006). Even though micro-level agents' behavior tends to cause the system to self-organize and converge, certain small events (disturbances) can have big global impacts and trigger large relaxation events that may encompass any number of agents (e.g., stock market crashes). It is the so called critical states or the self-organized criticality, which ought to be understood in order to be able to predict and prevent.

An overview of the fundamentals of CAS, as discussed above, is illustrated in Figure 4.

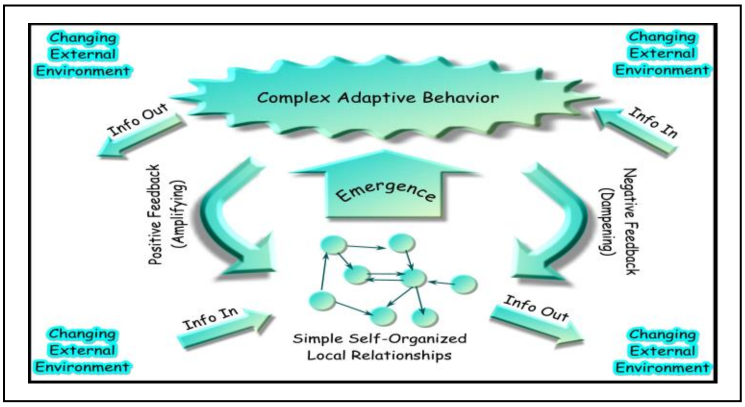


Figure 4. An Overview of the Fundamentals of CAS (Source: http://en.wikipedia.org/wiki/File:Complex-adaptive-system.jpg)

A complex adaptive system is intriguing in that the emergence of macroscopic order, while being a logical consequence of agents' actions and interactions in a given environment, are often unintended and unforeseeable. Therefore, one central aim of CAS study is to uncover how systems of interacting agents can lead to emergent phenomena. As a relatively new stream of thought, CAS theory is still accumulating insights regarding emergence in specific contexts (Nan, 2011). Although there is no universally agreed upon paradigm describing CAS (Gell-Mann, 1994), three elements have been consistently recognized as the core of the theory: *agents*, *interactions*, and *environment*. In the following subsection 2.3.2, the three elements will be discussed together with the agent-based modeling methodology.

2.3.2 Agent-based Modeling (ABM)

> "The use of computational, especially agent-based, models has already shown its value in illuminating the study of economics and other social processes."
>
> --Kenneth J. Arrow, Winner of the Nobel Prize in Economics

Indeed, over the past two decades, agent-based modeling (ABM) has become a much more widely accepted methodology for studying CAS. In this approach, complex systems are modeled as collections of autonomous interacting entities (agents) that operate in a computational world with encapsulated functionality. ABM aims to re-create or predict the emergence of complex phenomena through simulating the actions and interactions of multiple autonomous agents in a given environment, in order to assess the effects on the system as a whole. As such, not only real-world systems can be *simulated* with verisimilitude; they can also be efficiently and robustly *designed* and *constructed* on the basis of ABM principles (Borrill and Tesfatsion, 2011).

As discussed earlier, CAS exhibits two major properties: (1) consisting of heterogeneous interacting agents; (2) exhibiting emergent behavior, which arises from the interactions of the agents, and cannot be deduced simply by aggregating the properties of the agents. When the interaction of the agents is contingent on past experience, and especially when the agents continually adapt to that experience, mathematical analysis is typically very limited in its ability to derive the dynamic consequences (Axelrod and Tesfatsion, 2006). In this sense, ABM is an alternative way of analysis.

ABM starts with assumptions about agents, their interactions, and the environment which the agents operate in and interact with, and grows them in an artificial environment, as explained in the following.

Agents

Agents can represent individual actors (e.g., consumers, sellers, or voters), social entities (e.g., families, firms, communities, government agencies, and nations), or institutions (e.g., markets and regulatory systems). This type of agents is called data-gathering decision makers. Agents can also represent biological entities such as crops, livestock, and forests, as well as physical entities such as infrastructure, weather, and geographical regions. This type of agents is called passive world features. Data-gathering agents are typically assumed to have *bounded rationality*, meaning that agents' adaptive behavior (say decision making) is limited by the information they have, the cognitive and time constraints, and by what they perceive as their own interests such as economic benefits, social status, or survival needs (Simon, 1957).

In computational settings, the key properties of agents include: autonomy (capable of operating and making decisions on its own), sociability (capable of interacting with other agents), reactivity (capable of responding to a change of environment), proactivity (capable of acting on its own initiative in order to achieve certain goals/utilities), and adaptivity (with sophisticated learning capabilities) (Wooldridge and Jennings, 1995).

Furthermore, each agent consists of attributes and behavioral rules. Attributes are the internal states of agents. They can be fixed (e.g., consumer's price sensitivity) or varied (e.g., number of family members) over time (Epstein and Axtell, 1996). Variation in attributes makes agents distinguishable, thereby enabling agents to selectively respond to the behaviors of other agents (Holland, 1995). On the other hand, similarity among attributes allows agents to be grouped into classes. In addition, attributes are important parameters of behavioral rules, as described below.

Behavioral rules are the schemata governing an agent's attributes and actions. They are in the form of a set of input/output statements linking an agent's perception of the surrounding environment and other agent's behaviors to changes in its own internal states or actions (Drazin and Sandelands, 1992; Epstein and Axtell, 1996; Holland, 1995). Behavioral rules usually implement in the following sequences. Initially, a given agent obtains the perception of the environment and the behaviors of other agents. The breadth and depth of the perception are bounded by the given agent's attributes such as goals, utilities, and vision (Miller and Page, 2007). Next, based on the information collected from the first step as input, a set of statements (usually in IF/THEN syntax) is executed, which may either maintain or change the given agent's attributes or actions. These statements are decision-making rules, ranging from simple-fixed heuristics (e.g., 'tit for tat' in a cooperation game; Axelrod, 1997) to elaborate optimization routines that evolve over time (e.g., checkers-playing program; Samuel, 1959). Then, output from the second step solicits feedback from other agents and/or the environment, which in turn will either positively or negatively affect the behavior of the given agent. Consequently, the given agent adapts/learns from current experience by combining the elements of previously successful behavioral rules (Holland, 1995). As such, behavioral rules encapsulate the functionality for agents to evolve in response to information and feedback from other agents and the environment. Additionally, behavioral rules lead to another core element of CAS when unfolding in time and space: interactions (Drazin and Sandelands, 1992).

Interactions

Interactions reflect the mutually adaptive behaviors of agents and are the most commonly observed structures in a CAS (Drazin and Sandelands, 1992). Many studies have explicated interactions through diverse examples such as spontaneous coordination in standing ovations (Miller and Page, 2007), bilateral price negotiation in trading markets (Epstein and Axtell, 1996), and cooperation in groups (Arrow et al., 2000).

To simplify, interactions can be viewed as a function of agents, behavior rules, and connections. Agents are the bearers of interactions. As discussed above, behavioral rules contain the logic for agents to change their attributes and actions. Recurrent applications of behavioral rules can generate a temporal stream of interaction patterns.

Connections are relational links among agents. An agent's attributes not only make it distinguishable, but also define with which agent to form relational links and which not to. As attributes change over the course of interactions, connections may also evolve. In a CAS, connections specify possible channels through which interactions can take place.

Environment

The environment is where agents operate and interact with (Epstein and Axtell, 1996). The environment could encompass almost all external structures, such as macro- or micro-economic situation, political makeup, and physical circumstances in nature, which provide important conditions for actions and interactions to unfold. Meanwhile, external structures can be modified by the ongoing actions and interactions of agents. Hence, agents and the given environment surrounding them are in a mutual-influence relation. A good example is the artificial society model by Epstein and Axtell (1996). In this model, agents move towards resource-rich locations in an environment. While the agents collect resources from the environment, their resource-collecting behavior causes the redistribution of resources among different resource-bearing sites. Depending on the research question at hand, the environment can be represented as an agent in order to model how it will affect the agent's behaviors and the system at large.

In short, by simulating the three basic components of a CAS (i.e., agents, their interactions, and the environment), ABM allows to generate real world like scenarios in laboratory settings, so as to gain empirical understanding about the dynamics of a complex system.

The merits of ABM in studying CAS can be summarized as follows.

Firstly, for systems that are strongly interactive, it is often not practical (or even possible) to predict their global outcomes in advance of actual implementation even when their laws of motion are known (Borrill and Tesfatsion, 2011). ABM allows one to examine how agents' interactions with each other and with the environment can collaboratively create emergent outcome in a controlled manner.

Secondly, ABM, in particular, allows one to investigate whether and how the proposed changes in behavior rules (e.g., pricing strategy) or environment (e.g., economic policy) will generate new emergent behavior or adverse unintended consequences through simulation, in order to evaluate the possible impacts of such proposed changes on the real world. In so doing, it also aids in designing socially desirable strategies, policies, processes, and institutions efficiently.

Electricity Retail Market as a CAS

In relation to the current research, the electricity retail market is considered as a CAS. The electricity market participants (agents) include electricity users (consumers), electricity suppliers (retailers), load serving entities (such as transmission and distribution system operators), and electricity producers (see the terminology concerning the electricity industry in Chapter 3). The underlying mechanism is the economic principle of demand supply equilibrium, in addition to the physical laws governing electricity generation and transmission.

The implementation of bi-directional communication smart meters (AMI) makes demand response possible. Under demand response scenario, electricity users have the option to adjust their electricity consumption based on real-time price signals. The adjusted electricity usage will result in changes in aggregated hourly demand curves. Accordingly, the electricity suppliers (retailers) could modify both their pricing strategies in the retail market and their bidding strategies in the electricity wholesale market based on their economic objectives. On the wholesale side, the electricity price formation is largely dependent on the electricity producers' generation capacity, the marginal costs of electricity generation, demand variations, and transmission system capacity. The consumption adjustment by the electricity consumers will have a ripple effect on trading activities in the wholesale market. Apart from the above mentioned dynamic interactions between consumers, electricity suppliers, and/or producers, the external factors (i.e., environment) such as macro-economic situation, fuel prices, the weather, and intermittent energy sources (e.g., wind and solar), etc. will also have impacts on the electricity demand and supply activities.

In designing dynamic pricing or demand response programs, the possible impacts of new pricing strategies on electricity demand, especially in advance of actual implementation, need to be assessed. Therefore, it is important to simulate the interactions between the electricity supplier's pricing and electricity consumers' consumption adjustment in order to capture and understand the emergent outcome.

Motivated by the merits of ABM in studying emergence of CAS, this research is to investigate the potentials of ABM in price-responsive demand simulation (see Chapter 6). The ABM-based price-responsive demand simulation is intended to function as a prescriptive BI solution for the electricity retail market. From the point of view of the entire research design, ABM simulation is conducted as a proof-of-concept.

2.4 Chapter Summary

This chapter specified the methodological basis of the research. In light of design science research (DSR) paradigm in the IS field, the entire research is constructed through steps of (1) problem identification and motivation, (2) define solution objectives, (3) design and development, (4) demonstration, (5) evaluation, and (6) communication, with relevance-, rigor-, and design- cycles iterations. Motivated by the notion of knowledge discovery (KDD) through data mining, the Self-Organizing Map (SOM) is selected as visual data mining method for electricity consumption behavior profiling and electricity distribution regulation benchmarking. Meanwhile, the tenets of complex adaptive systems (CAS) are employed to propose a theoretical lens and a computational simulation tool (i.e., ABM) particularly suitable for capturing the emergence of demand response electricity retail market. In the next chapter, the terminology presented so far concerning the electricity markets will be detailed, and the unfolding influential events in the European electricity sector which are relevant to this research will be pinpointed.

Chapter 3

State of the Art in the European Electricity Industry

Electricity supply is regarded as one of a nation's top security issues worldwide. Electricity supply has become the backbone and infrastructure of a modern economy. Yet, for most people living in the post-industrial and information-intensive society, electricity is just a necessity in our daily life. In fact, the electricity supply business has gone through considerable changes during the recent decades, especially into the 21st century with the development of renewable energy sources (RES), and technological innovations in power generation and distribution. This chapter will outline influential topics in the electricity industry by introducing basic concepts such as *electricity markets*, *smart metering*, *smart grid*, *demand response*, etc. Related research regarding data mining in the energy sector and agent-based modeling for electricity market analysis will be highlighted. The purpose is to give an overview of the problem domain, and put the current research into perspective.

3.1 Deregulated Electricity Markets

This section will provide a brief review of electricity market deregulation, known as *restructuring* and *unbundling*, with a particular focus on the European electricity industry. The resulting value chain of the European electricity industry and the terminology in question will also be presented.

3.1.1 Restructuring and Unbundling

The power sectors around the world continue to evolve from regulated, vertically integrated monopoly structures, to open markets. The attempt is to promote and maintain competition in electricity generation and supply functions, to ensure a greater choice for consumers, and thus, to impose pressure on energy companies to offer the best possible value and services. This process is known as *electricity market deregulation*, or by other alternative terms such as *liberalization*, *restructuring* and *unbundling*.

Electricity market deregulation has been implemented by different countries as early as the 1980's. In different deregulation processes the institutions and market designs often varied, but the underlying concepts were similar. The heart of many of these reforms is to unbundle the operations of electricity generation, transmission, distribution, and supply, and to establish a wholesale electricity market and a retail electricity market. The purpose of so doing is to introduce competition and market mechanisms into electricity generation and supply, while the activities of transmission and distribution are still subject to regulation as they are considered to be natural monopoly.

For instance, in 1991, Norway was the first country in Europe that started to liberalize its electricity market, followed by Finland in 1995 and Sweden in 1996, respectively. Consequently, Norway, Finland, and Sweden had completed full liberalization as of 1995, 1997, and 2000, respectively (Jamasb and Pollitt, 2000). This means that the electricity markets in the three countries opened access to all users, and the electricity generation and retail became competitive business fields, whereas the transmission and distribution businesses remained under regulation. As a result, the Nordic countries (Norway, Sweden, Finland, and Denmark) had established a common electricity market Nord Pool for wholesale electric power trading by 2000, which is known nowadays as Nord Pool Spot. Nord Pool Spot is owned by the Nordic national grid companies: Statnett, Svenska Kraftnät, Fingrid, and Energinet.dk, and the Baltic transmission system operators: Elering, Litgrid, and Augstsprieguma tikls (AST). It operates in Norway, Sweden, Finland, Denmark, Estonia, Latvia, and Lithuania, serving both the day-ahead market (known as Elspot) and intraday market (known as Elbas). It was the world's first multinational exchange for trading electric power, and is the leading electricity wholesale market in Europe in terms of volume traded (493 TWh in 2013) and market share (84% of total consumption of power in the Nordic/Baltic market in 2013).

To harmonize the fully-opened internal electricity market within its member states, the European Union (EU) has enacted a number of legislations (e.g., directives 96/92/EC, 2003/54/EC, and the so-called Third Energy Package (2009) including two directives and three regulations). The aim of these legislations is to break down national barriers to power trade, allowing energy flows to be determined by supply and demand rather than local rules, thus improving choice and services for consumers, and overall security of supply. Apart from requiring the functional separation (unbundling) of generation, retail, transmission, and distribution of vertically integrated energy companies (but not necessarily in ownership unbundling), the latest Third Energy Package also addresses the legal separation of generation and transmission with three options: ownership unbundling, independent system operator (ISO), or independent transmission operator (ITO).

3.1.2 Electricity Value Chain

Following the EU's initiatives, the European electricity sector has evolved to become uniform, and its value chain is illustrated in Figure 5.

The market participants in the electricity markets include producers, transmission system operators (TSOs), distribution system operators (DSOs), suppliers, and consumers. The producers, TSOs, DSOs, and the suppliers consist of the supply side of the electricity value chain, versus the electricity consumers on the demand side. The power generated by electricity producers is transported to the consumers, either via the transmission grid and the distribution grid, or directly using an exclusive cable. The transmission grid is generally a national level network used to transport bulk electric power at high voltage from production sites to distribution grids. Accordingly, the distribution grid is a local level network delivering medium to low voltage electric power to electricity consumers. TSOs and DSOs are the entities in charge of running, maintaining, and developing of the respective grids. TSOs are responsible for coordinating the dispatch of generating units to meet the expected demand of the system across the transmission grid. TSOs provide grid access to the electricity market participants according to non-discriminatory and transparent rules. In order to ensure the security of supply, they also guarantee the safe operation through balancing and ancillary services, and maintenance of the system. In many countries, TSOs are in charge of the development of the grid infrastructure too. TSOs in the EU internal electricity market are entities operating independently of the other electricity market participants. Many EU member states only have one TSO, for instance, Fingrid in Finland, Réseau de Transport d'Electricité in France, while e.g., Germany and the UK have four TSOs, each. In addition to running the distribution networks, DSOs also take care of billing and meter-reading, which means that the DSOs will play an important role in the unfolding smart meter rollout, as will be explained in Section 3.2. The electricity suppliers (retailers) are responsible for trading electricity on a purely commercial basis, which involves participating in auctions on the electricity wholesale market and entering retail contracts with the consumers.

The payment flows in relation to electricity supply are as follows: on the one hand, the consumers buy the power from various suppliers; on the other, the TSOs and DSOs will charge network tariffs to cover the cost of running and maintaining the grids. As such, the electricity retail rate which the consumers are facing consists of energy price and network tariff. Energy price is usually derived from the wholesale power price as described below, while network tariff is based on local DSO's operational efficiency (see Chapter 7 for details).

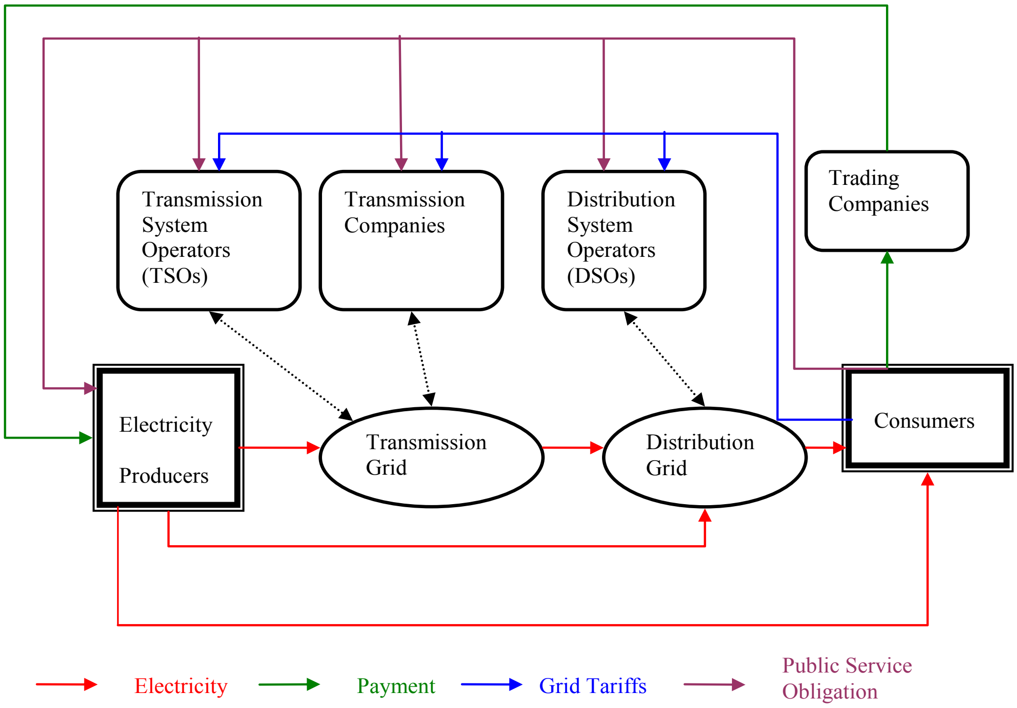


Figure 5. The Electricity Value Chain in Europe

3.1.3 Some Concepts in Electricity Markets

The restructured electric power sector can be divided into *electricity wholesale* and *retail* markets. As illustrated in Figure 6, the generation (producers) and transmission (TSOs) operations constitute the wholesale side, while the distribution (DSOs) and consumption (consumers) form the retail side. As mentioned earlier, the suppliers partake in both wholesale and retail markets, as do other load serving entities such as load aggregators.

The ongoing integration of the European internal electricity market has resulted in seven regional wholesale electricity markets:

- Central Western Europe (Austria, Belgium, Germany, France, the Netherlands, Switzerland);
- British Isles (UK, Ireland);
- *Northern Europe (Denmark, Estonia, Finland, Lithuania, Norway, Sweden);*
- Apennine Peninsula (Italy);
- Iberian Peninsula (Spain and Portugal);
- Central Eastern Europe (Czech Republic, Hungary, Poland, Romania, Slovakia, Slovenia);
- South Eastern Europe (Greece).

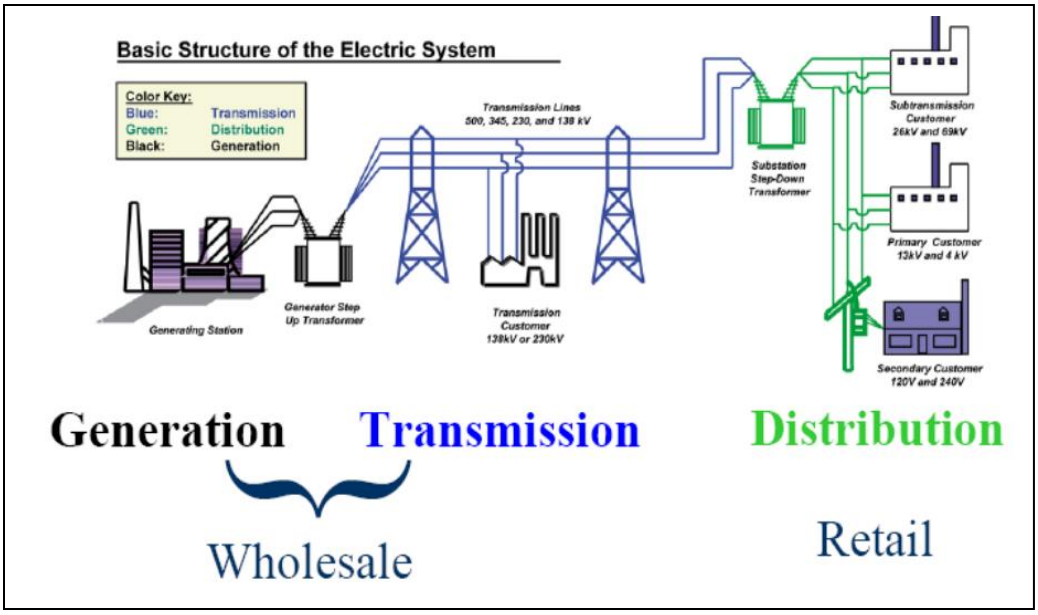


Figure 6. Wholesale and Retail Power System Operations
(Source: http://www.nerc.com)

The role of the wholesale market is to allow trading between energy producers, energy intensive industries, large consumers, distributors, utility companies, retailers, and financial institutions both for short-term (usually day-ahead and intraday) delivery of electricity and for future delivery periods (i.e., derivatives).

As the empirical study is from the viewpoint of the electricity industry in Finland and Estonia, the following discussion regarding wholesale market mechanism is mainly based on the practices of the Northern Europe electricity wholesale market to which Finland and Estonia belong, specifically Nord Pool Spot. As mentioned earlier, the Nordic electricity wholesale market (Nord Pool Spot) operates a day-ahead market (Elspot) and an intraday market (Elbas).

The *day-ahead market* is where the buyers and sellers submit their bids and offers for the power which will be delivered the *following* day. All purchase and sell orders are aggregated into two curves for each delivery hour – an aggregate demand curve and an aggregate supply curve. The *system price* for each hour is determined by the intersection of the aggregate supply and demand curves, which are representing all bids and offers for the entire Nordic region. In other words, the spot market price is set where the curves for sell price and buy price meet. The system price is the market clearing price that disregards the available transmission capacity between bidding area[3] in the spot market, and is very often used as the reference price for the financial market.

While supply and demand are the key factors determining the hourly spot market prices, *transmission capacity* also plays a role. Transmission capacity is one form of physical constraint embedded in the electric power system. It defines how much power can be transported through the grid. Bottlenecks in the transmission system can occur if large volumes need to be transmitted to meet demand. To relieve grid congestion, different *area prices* are introduced. If the need for transmission exceeds the available transmission capacity, the prices are decreased in surplus areas and increased in deficit areas, which results in different area prices.

The majority of the volume handled by Nord Pool Spot is traded on the day-ahead market. For the most part, the balance between supply and demand is secured. However, incidents may take place between the closing of Elspot at noon CET and the physical delivery starting the next day. An *intraday market* is necessary in order to offset the imbalance between day-ahead contracts and actual power production and delivery. In the intraday market (Elbas), buyers and sellers can trade volumes close to real time to bring the market back in balance. Elbas is a continuous market, and trading takes place every day around the clock until one hour before delivery. Prices are set based on a first-come, first-served principle, where best prices – highest buy price and lowest sell price come first.

Daily power trading in the wholesale market is driven by the market participants' planning. A buyer, typically a retailer, needs to estimate how much energy volume will be needed to meet demand in the next day, and how much the company is willing to pay for this volume, hour by hour. The seller, for example, a hydroelectric power plant, needs to assess how much energy it can deliver and at what price, hour by hour. However, the information about market conditions, communicated by the wholesale prices, is not conveyed to consumers without automatic meter reading, i.e., the majority of households and small to medium sized businesses. Historically, these consumer segments could only choose between fixed retail rates for certain periods of time. Since these consumers are not aware of the varying costs of electricity generation and transmission reflected in the wholesale prices, they have no incentives to

---

[3] For each Nordic country, the local TSO decides, how many bidding areas the country is divided into. Today there are five bidding areas in Norway. Eastern Denmark and Western Denmark are always treated as two different bidding areas. Finland, Estonia, Lithuania and Latvia have one bidding area of each. Sweden was divided into four bidding areas on November 1st, 2011.

adjust their consumption, which results in low demand elasticity. Their retailers hence must submit price insensitive bids in the wholesale market, and are forced to pay any price in order to serve the demand. This situation indicates the disconnection between the wholesale and the retail markets. As a consequence, this disconnection has contributed to inefficient allocation of production and transmission resources in the short term, non-optimal investments in capacity in the long run, higher price volatility, and an electricity system easily being exposed to exercises of market power. In an attempt to improve the functioning of electricity markets, the deployment of smart metering and smart grids has been initiated within the EU, as will be presented in the next section.

3.2 Smart Metering and Smart Grids

3.2.1 Smart Metering

Intelligent metering systems are promoted for several reasons in the EU Third Energy Package: firstly with the aim to promote energy efficiency and demand side management measures; and secondly with the aim to ensure active participation of consumers in the energy market. The implementation of these metering systems may be subject to an economic assessment. Furthermore, the Electricity Directive (2009/72/EC) states that, subject to that assessment, member states should prepare a timetable with a target of up to 10 years for the implementation of intelligent metering systems. Where roll-out of smart meters is found feasible, at least 80 % of consumers should be equipped with intelligent metering systems by 2020.

Due to differing regulations on the national level, adoption of smart metering varies greatly within the EU, as shown in Figure 7. By the beginning of 2013, two countries have completed their roll-out (Sweden 100% and Italy 95% of customers); one country (Finland) should reach the 80% target by the end of 2013; and fifteen countries are rolling out or plan to roll out smart meters. Eleven of these will target 95% or more; three countries will target 80%; and one (Germany) will target 15%. Regarding the financing of the roll-out, a majority of countries (12) is financing through DSOs, except in the UK where it is financed by suppliers.

The technical design of smart metering systems varies widely across countries. In order to define a smart electricity meter, the following functionalities are associated:

- remote reading;

- two-way communication;
- interval metering;
- remote management;
- home automation;
- web portal.

Figure 7. EU Rollout of Smart Meters (Source: European Commission Joint Research Centre (JRC) – Institute for Energy and Transport)

Even though roll-outs have already been completed in some countries, most have not yet begun or are still in the early stages. Therefore, it is premature to project the impacts. However, some application scenarios can be expected from the technical design of smart metering systems, such as demand response, remote power capacity reduction, as well as remote activation and de-activation of supply. In addition, smart metering systems ought to be featured by the following (CEER, Ref: C13-RMF-54-05):

- an alarm alerting the customer of exceptional energy use;
- an open gateway through which the customer can access and control their consumption;
- a remote upgrade capability;
- the capability to measure injected as well as consumed energy;
- the capability to receive immediate information on non-notified energy interruptions at the connection point.

3.2.2 Smart Grids

Power transmission and distribution networks are among the largest and most complex physical man-made systems. They have gone through technological and structural revolutions from being formed as rather small, monopolistic, and fully controllable systems to being very large interconnected systems, with a large share of distributed and unpredictable renewable energy sources (RES), such as wind and solar. The systems are deregulated and are hence increasingly ruled by a free real-time electricity market. Addressing the new challenges of transporting electricity is by means of the conceptual invention of a smart grid.

The EU's 20-20-20 goals (20% increase in energy efficiency, 20% reduction of CO2 emissions, and 20% renewables share by 2020) stipulate that today's electricity grids need to become smarter, more efficient and customer-oriented than ever before, to not only provide power but also information and intelligence.

Although there is no standard global definition, the EU's smart grids Technology Platform defines smart grids as "electricity networks that can intelligently integrate the actions of all users connected to it - generators, consumers, and those that do both – in order to efficiently deliver sustainable, economic, and secure electricity supplies." (European Technology Platform SmartGrids, 2010).

In essence, smart grids are to leverage state-of-the-art ICT technologies – distributed sensors, real-time communication networks, advanced control systems, networked embedded systems, high level automation – to cope with uncertainty in supply, demand, and pricing of electricity.

The potential benefits of smart grids include:

- optimizing grid operation and usage (e.g., reducing network losses) and grid infrastructure;

- facilitating higher penetration of renewable energy sources (e.g., wind) and distributed generation (DG) such as small windmill or rooftop solar cells;
- helping consumers better participate in the market not only by using their energy more efficiently (e.g., through smart metering) but also by allowing consumers to act also as producers selling back their excess electricity (e.g., CHP (combined heat and power) or plug-in electric vehicles);
- maintaining and improving the existing network services efficiently (e.g., adequate short circuit power at point of grid connection, efficient and reliable alarm and fault management for self-healing procedures in distribution networks, adequate (bidirectional) protection concepts for distributed generation, etc).

Even though it may seem that the usage of the terms *smart metering* and *smart grids* is not distinct in the publications connected to this research, they are actually different. Very often smart metering is considered as an essential step in the transition towards a smart grid. The conceptualization of smart metering and smart grids is illustrated in Figure 8. It implies that a smart grid ought to integrate all the electric elements together. It encompasses the conventional elements such as centralized generation, transmission and distribution networks, inter-connections, as well as the innovative and unfolding elements such as distributed generation, electric vehicles, and smart metering. With respect to smart metering, apart from the physical elements of meters and displays, residential demand and customer behavior play central roles. These are what the current research focuses on – consumption behavior profiling and demand simulation.

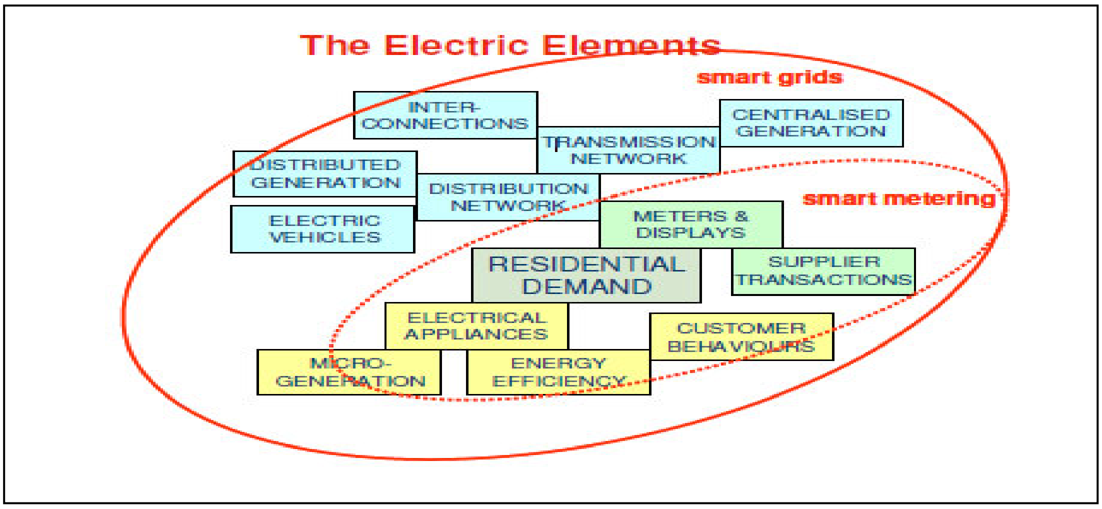


Figure 8. Conceptualizing Smart Grids and Smart Metering
(Source: CEER FS-10-01)

3.3 Demand Response

With a large number of smart meters being deployed through various smart grid initiatives, the utilities today are anxious to know how to utilize the wealth of data coming out on a sub-hourly basis. Demand response has often been mentioned.

It is argued that increasing demand response is important to achieving a well-functioning, efficient, and reliable electricity market (Ericson, 2007; Faruqui and Wood, 2008). So, what is demand response? According to the definition adopted by CEER (Council of European Energy Regulators) in the context of the retail market, *demand response* refers to

> 'changes in electric usage by end-use customers/micro generators from their current/normal consumption/injection patterns in response to changes in the prices of electricity over time, or to incentive payment designed to adjust electricity usage at times of high wholesale market prices or when system reliability is jeopardized. This change in electricity usage can impact the spot market prices directly as well as over time.'

This definition emphasizes the innovative pricing aspect of demand response, as direct load control is an alternative method for achieving demand response goals. Hence, the discussion regarding demand response throughout this study follows the CEER's definition, i.e., focused on the perspective of innovative pricing.

To differentiate demand response from demand side management, Figure 9 describes the full picture of demand side management, where demand response is one dimension, based on CEER's categorization. The dimensions of energy response and direct load control are beyond the scope of this thesis.

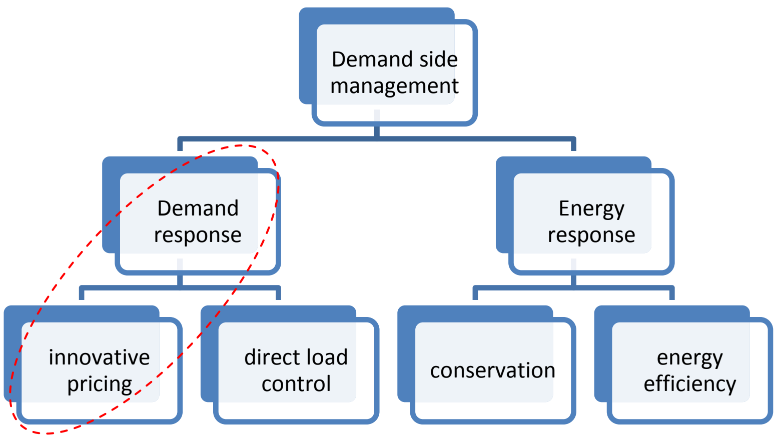


Figure 9. Demand Side Management Diagram

It is believed that deploying demand response will shift the paradigm of electricity markets in many ways. Foreseeably, consumers will be able to manage and adjust their electricity consumption in response to real-time information and changing price signals. Accordingly, electric utilities will be capable of altering the timing, level of instantaneous demand, or the total electricity consumption at times of high wholesale market prices or when electric system reliability is jeopardized (Albadi and El-Saadany, 2007). Such a price-responsive interaction between demand and supply will in turn impact the electricity production planning, production portfolio management, and the investment in generation and transmission capacities, and eventually, improve the link between wholesale and retail electricity markets, which to a great extent are currently disconnected.

The potential benefits of full participation by demand include flattening daily load patterns, optimizing the production portfolio by mitigating the variability of generation from renewable sources, and reducing the investment in reserve capacity needed to maintain resource adequacy and system reliability (Schuler, 2012), thus improving overall market efficiency.

Figure 10 summarizes the future scenario of electricity markets with smart metering, smart grids, and demand response in place.

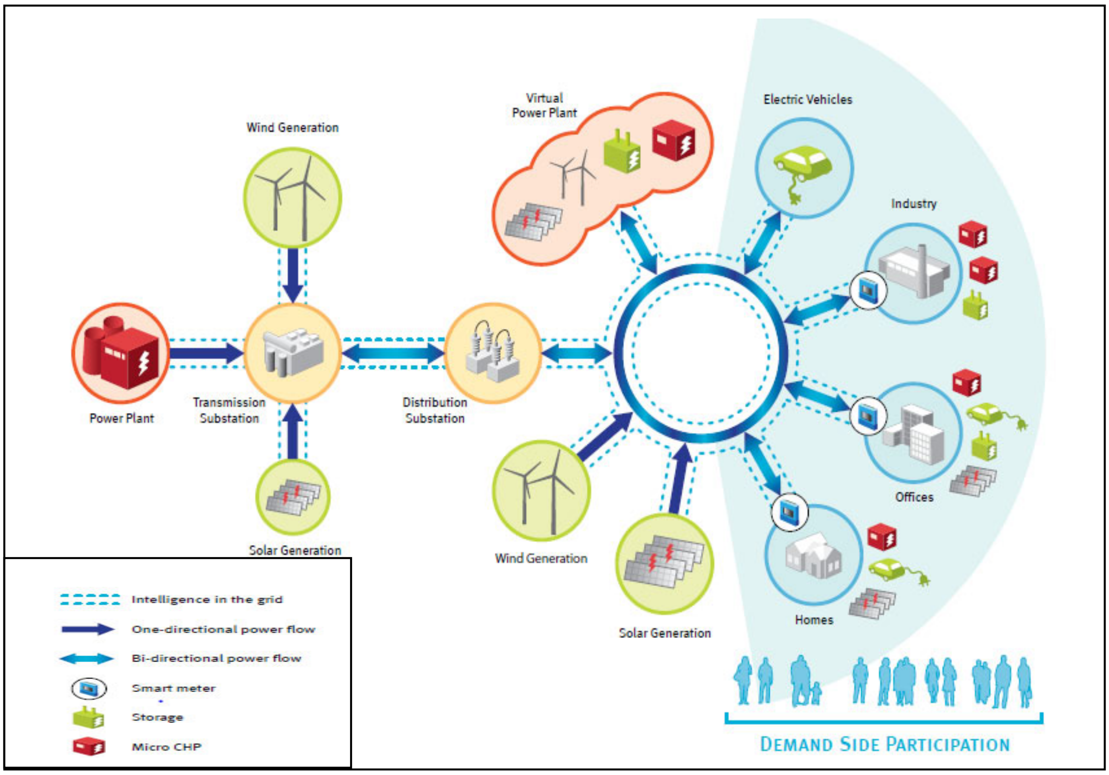


Figure 10. Electricity Markets of Tomorrow
(Source: EURELECTRIC—10 Steps to Smart Grids)

3.4 Relevance to the IS Field and Related Research

This section presents the relevance of IS studies to the electricity industry by outlining related research. The aim is to highlight the two research methods, i.e., data mining and agent-based modeling in the problem domain, which are employed in this thesis.

3.4.1 Data Mining in the Energy Sector

As discussed in Chapter 2, data mining is a computational process of discovering useful patterns in large data sets. Data mining has been widely examined in power system and energy markets studies. For instance, in Ramos and Liu (2011), data mining has been applied for power system stability assessment, transmission expansion planning, viscosity control in fossil fuel power plants, generation reliability in power markets, short-term load forecasting, and control of wind generators. It has also been used to study learning behavior of electricity markets players, electricity market pricing, bids in electricity markets, as well as profits in electricity markets.

A number of data mining techniques have been investigated in various studies, in addition to the SOM method stated in Section 2.2.2. The specific data mining techniques employed range from neural networks (Verdu et al., 2004), reinforcement learning (Ernst et al., 2004), *k*-means (Ferreira et al., 2011), to decision trees (Figueiredo et al., 2005), and support vector machines (Ekici, 2009). For instance, Rodriguez and Anders (2004) introduced a method for forecasting energy prices in the Ontario competitive power system market using neural networks, fuzzy logic, and a combination of the two. Vale et al. (2011) presented a method to mine a large set of operation scenarios in order to support system operators and/or Virtual Power Players (VPPs) to determine effective and efficient demand response programs.

In particular, a variety of research studies have addressed classification and clustering issues in load profiling, such as Pitt and Kitschen (1999), Espinoza et al. (2005), Figueiredo et al. (2005), Nizar et al. (2006), Romas and Vale (2008), and so forth. On the other hand, electric utilities can use the revealed consumption patterns to improve their business processes. For instance, Chicco et al. (2003) use load profiles to study the margin opportunity of a distribution company in the design of optimized tariffs to identified customer classes. Espinoza et al. (2005) develop forecasting models to identify typical daily customer profiles in cooperation with the Belgian National Grid Operator ELIA.

Given the increasing adaption of smart metering and smart grids technologies, it is argued that data mining techniques are of central importance in the transformation of power systems and electricity markets (Keshav and Rosenberg, 2011; Ramos and Liu, 2011). It is certain that the timely measured smart meter data which are streaming in on a sub-hourly basis is posing a continuous challenge to data mining research.

3.4.2 Agent-based Modeling in Electricity Market Analysis

Agent-based modeling and simulation has become a promising and very active research area for electricity market modeling. The perspective of this research methodology is to tackle the complexity of electricity market mechanisms. With the electricity industry deregulation and restructuring, an increasing number of studies have appeared in the past decade.

A great deal of research in the field of agent-based modeling of electricity markets has concentrated on the analysis of market power and market design in wholesale electricity trading. Various wholesale electricity market simulation models were developed, for instance, by Bower and Bunn (2000) for the England and Wales electricity market, Bower et al. (2001) for the German electricity sector, Cau and Anderson (2002) for the Australian National Electricity Market, and a research group at Iowa State University (Koesrindartoto et al., 2005; Sun and Tesfatsion, 2007) for the U.S. Wholesale Power Market Platform.

In addition, different computational algorithms have been examined for agent-based electricity market modeling, including genetic algorithms for representing the agents' bidding behavior (Nicolaisen et al., 2000; Richter and Sheblé, 1998), Erev-Roth reinforcement learning algorithm (Nicolaisen et al., 2001; Petrov and Sheblé, 2001), and rule-based learning mechanisms combining reinforcement learning and genetic algorithms (Bagnall and Smith, 2005).

An alternative body of agent-based research modeled electricity consumer behavior at the retail level. Zhou et al. (2011) studied the consumption behavior of commercial buildings with different levels of demand response penetration in different market structures. Ehlen et al. (2007) presented a simulation based on the N-ABLE™ model, in which they studied the effects of residential real-time pricing contracts on demand aggregators' load, pricing, and profitability. Müller et al. (2007) investigated the interdependencies between the customer's engagement and the suppliers' pricing strategies in the German retail market. Additionally, some agent-based studies focused on

Time of Use (TOU) pricing for residential customers in different contexts (Roop and Fathelrahman, 2003; Hämäläinen et al., 2000).

In recent years, there is a variety of agent-based electricity market research being published in relation to the emergence of smart metering, distributed generation (DG), and demand response, such as Guerci et al., 2010; Li and Tesfatsion, 2009; Thomas et al., 2012; just to name a few. Among them, Cai et al. (2011) simulated distribution systems with high penetration of photovoltaic generation using ABM. Kok et al. (2009) studied intelligent distributed coordination for embedding renewables and distributed generation (DG) using multi-agent systems (MAS). Zhang and Nuggall (2011) evaluated governmental policies on promoting smart metering diffusion in electricity retail markets via agent-based simulation.

The above-mentioned publications present the heterogeneity of research topics, modeling assumptions, and computational techniques with respect to ABM for electricity market analysis. The increasing body of research shows that ABM is a valid research methodology regarding electricity market modeling.

3.5 Chapter Summary

This chapter presented the state-of-the-art in the European electricity industry, by characterizing *restructuring and unbundling*, *smart metering*, *smart grids*, and *demand response*. During this course, electricity market mechanisms, legislative and regulatory aspects, and influential and unfolding phenomena were introduced. In addition, the existing gap between electricity wholesale and retail markets, and the obstacles for achieving market efficiency, were discussed. Finally, the relevance of IS research in the problem domain was highlighted through a literature review. The purpose of so doing is to put this thesis into perspective: advancing BI in facilitating the understanding of customer behavior and residential demand response for the electricity markets of tomorrow. The remaining chapters, from Chapter 4 to 7, will concentrate on the four building blocks of BI, which have been investigated in this study:

- descriptive analytics based on visual data mining for consumption profiling in Chapter 4;
- holistic real-time BI enterprise architecture in Chapter 5;
- prescriptive analytics based on agent-based modeling and simulation in Chapter 6;
- a visualization application for electricity distribution benchmarking in Chapter 7.

Chapter 4

Consumption Profiling for Pricing Decision-Making Support

This chapter summarizes the research work published in Papers I & II. In the wake of the large scale smart meter roll-out to be completed in Finland by 2013, a business intelligence (BI) approach, namely consumption profiling, is proposed in this thesis for identifying new business opportunities enabled by smart metering. The said consumption profiling is based on visual clustering in the form of the Self-Organizing Map (SOM). Essentially, it is a descriptive analytics approach within the BI family. In the following, the empirical studies including the data, the modeling, and the results will be presented, and the implications will be discussed.

4.1 Empirical Studies

The empirical studies analyze quasi-daily smart meter data for approximately 12,000 customers in a Finnish region during 2007-2009. The objective is to study (1) what useful knowledge can be detected by such a visual data mining approach; and (2) what is the added value for the business practice in applying such an analytical method for decision-making support, especially with respect to pricing differentiation or designing demand response tariffs.

Data

The data investigated is from a case company's meter reading registers. For each meter, the electricity usage is registered with 27 hours 20 minutes time intervals, due to the automated meter reading system (AMR) and communication technology adopted (Turtle Automated Meter Reading system). The Turtle AMR uses the power line for data transmission (using power line to transmit data is very slow; a 48-bit message takes 27 hours to send.). The data is collected by a receiver installed at a substation and held until requested by a computer at the main office, then sent via SMS. Within each measuring interval (27h 20min), the Turtle AMR also calculates the highest rate of electricity usage of each meter (i.e., the Peak Load), with a particular timestamp (including actual date, hour, and minute). Therefore, the data from meter reading registers includes Meter ID, Electricity Usage, Reading Time, Peak Load, and Peak Time.

As mentioned above, each peak load is registered with a timestamp indicating the specific time when the peak appears (i.e., the Peak Time). The timestamp makes it possible to classify the peak load according to various time periods (e.g., 6-9, 9-16, 16-22, and 22-6) during data pre-processing.

In addition, the customer statistics from the case company's operational system have also been used in building the data-mining models. The customer statistics consist of Electricity Rate, Housing Type, Consumption Category, Fuse Type, etc.

Modeling

A set of modeling were carried out from different analytical perspectives. In Paper I, the customers' average consumption profiles during 2007-2009 were identified without regard to their conventional classification (i.e., customer categories and housing type). Then, the electricity contract choice was compared in light of their actual consumption patterns. In Paper II, customers' electricity consumption patterns in 2009 with a focus on time series analyses were studied.

As the empirical study in Paper II is closely associated with the discussion in the next chapter, the following part will be focused on presenting the modeling and the results reported in Paper II.

The modeling process was as follows: first, the SOM was used to cluster customers according to their electricity consumption similarity in 2009. Then, the consumption profile of each cluster was visualized through feature plane analysis. During the analysis, different variable sets in terms of day-of-the-week, season, and time band partition were compared in order to extract more detailed (hopefully unanticipated) information about the customers' consumption patterns.

In the data preparation phase, a great deal of data pre-processing work, including data integration, transformation, aggregation, and normalization, had to be performed to create customer signatures, with one record per customer and a range of variables capturing customers' demographic and consumption related features. We excluded the customers whose records included less than one year, or whose annual consumption was 0 kWh. There were in total 11,964 customers included in this study.

The variables used fall into two types based upon their purpose – one type is for describing the customer's general consumption and demographic profile, and the other is for investigating customers' weekday-weekend, seasonal, and time-band related consumption patterns. The variables are described in Table 1.

Table 1. Summary of Variables

Variables	Description
Average Consumption (kWh)	the customer's average consumption per 27hrs 20mins +/- 8mins
Average Peak Load (kW)	the customer's average peak demand, based on the highest load aggregated from three consecutive 20min intervals during each 27hrs 20mins period
Electricity Rate	the contractual electricity tariff, mainly 3 categorical attributes: Normal rate, Economic rate, and Time rate
Housing Type[4]	historical statistics, including 4 categorical attributes: Summer Cottage, Detached House, Town House, Multi-storeyed Building
Seasonal and day-of-the-week Consumption (kWh)	includes Weekday, Weekend, Jan.-Apr., May-Sep., Oct.-Dec., proposed: Winter (Oct.-Feb.), and Summer (March-Sep.) consumptions
Time-based Peak Load (kW)	the customer's average peak demand at various times of the day, including: Peak Load_Day, Peak Load_Night; proposed: Peak Load_6-9, Peak Load_9-16, Peak Load_16-22, and Peak Load_22-6

In this study, Viscovery SOMine v 5.0 was used to perform the visual data mining task. SOMine uses an expanding map size (meaning the ratio of X and Y dimensions, where X > Y) and the batch training algorithm, allowing for efficient training of maps (Deboeck and Kohonen, 1998). The SOM-Ward clustering method was also used to identify clusters based on actual consumption behavior, which eliminated the need for subjective identification of clusters (Vesanto and Alhoniemi, 2000). The seasonal and time-based variables were normalized according to their respective average values before map training, i.e., each entry in a variable field was divided by the mean of the entire field (Collica, 2007). The purpose is to address the relative significance of the value of a particular variable against the overall mean of that variable. In addition, all the variables included in the training process were scaled to comparable ranges in order to prevent variables with large values from dominating the result. Viscovery SOMine offers two forms of scaling, *linear* and *variance* scaling. Linear scaling is based upon the range of the variable, and is used as default when the range of the variable is greater than eight

[4] Categorical variables, such as Electricity Rate and Housing Type, must be split into binary dummy variables in order to be used with the SOM, as they represent nominal data with no inherent numerical order or distance.

times its standard deviation. Otherwise, variance scaling is used. In this study, range scaling was applied to the variables of Electricity Rate and Housing Type, while variance scaling was applied to the others.

During the training process, different combinations of parameters were examined, and the map was selected based on following criteria: average quantization error, normalized distortion measure, the meaningfulness of clusters, the visual interpretability, the smoothness of neighborhood of each node, and the SOM-Ward cluster indicator. The map was trained using a map size of 279 nodes, a map ratio of 100:49, and a tension (i.e., the radius of the neighborhood) of 0.5. In addition, the priority[5] of categorical variables such as Electricity Rate and Housing Type, as well as the proposed seasonal and time-based variables, was set to 0.

In order to evaluate the robustness of the training method, a supervised ten-fold cross-validation was conducted. The entire training dataset was firstly partitioned into 10 subsets, then using 9 out of the 10 subsets each time to reiterate the map training with the same set of training parameters as was described above. The map selecting criteria, as mentioned above, can be held over the ten-fold iteration.

Results

I. *<u>Cluster Profiles</u>*

The 11,964 customers are grouped into 4 clusters according to their consumption similarity in 2009. The results can be seen in Figures 11-13. Since the warm color code (e.g., red) in the SOM maps denotes high values while a cold color code (e.g., blue) represents low values, the characteristics of each cluster (I-IV) can be easily identified, as summarized in Table 2.

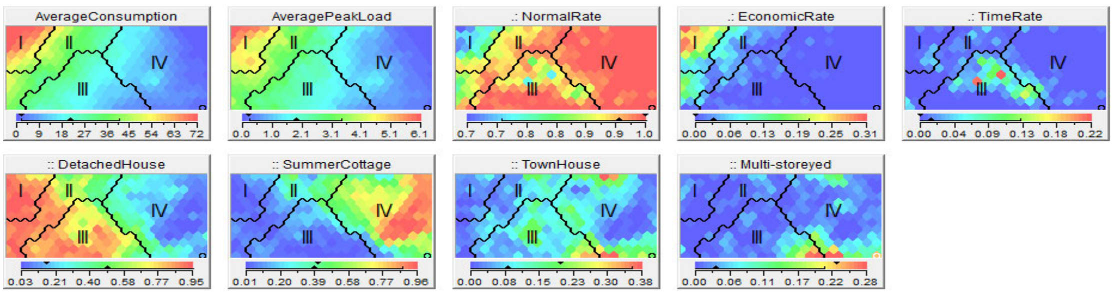


Figure 11. Cluster Profiles

---

[5] Priority is the weights of the variables. The higher the priority of a variable the higher is its influence on the map. Setting the priority of variables to '0' means that these variable have no influence on the training process, but their distribution in each of the segments can be visualized on the map for comparison and profiling purposes.

Table 2. Summary of Cluster Characteristics

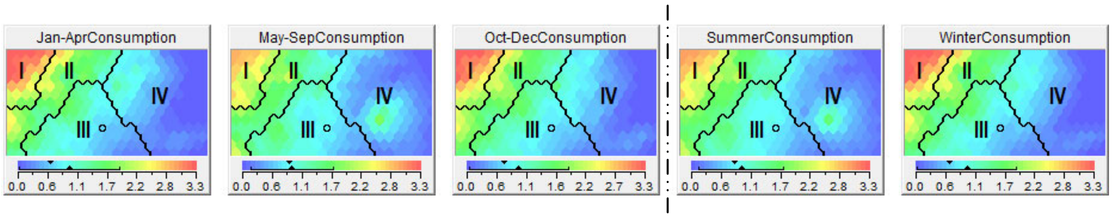
ID	Consumption Rank	Average Daily Consumption (kWh)	Average Peak Demand (kW)	Cluster Size and Percentage of Total Consumption (%)	Cluster Profile
I	High consumption	63.0	5.1	10.0, 28.9	Highest electricity demand; Highest proportion (19%) for Economic rate; The majority (88%) lives in detached house, while 7% in summer cottage, 4% in town house, and 1% in multi-storeyed building.
II	Medium-high consumption	39.3	3.2	17.0, 30.7	Medium to high electricity demand; The majority (94%) for Normal rate, while 5% for Economic rate; The proportion of summer cottage (18%) is the second highest after cluster IV, while the majority (75%) is in detached house, 6% in town house, and 1% in multi-storeyed building.
III	Medium-low consumption	20.3	2.0	25.9, 24.1	Medium to low electricity demand; Similar characteristics as those of cluster II, e.g., 96% for Normal rate and 75% in detached house; The proportion of town house is the highest (12%), while 9% in summer cottage and 4% in multi-storeyed building.
IV	Low consumption	7.5	0.6	47.1, 16.2	Lowest electricity demand; 99% for Normal rate; Highest proportion (70%) of summer cottage, while 18% is detached house, 8% town house, and 4% multi-storeyed building.

II. *Consumption Time Series Profiling*

The consumption pattern profiling was carried out with a focus on two types of consumption time series, including (i) consumption seasonality, and (ii) load patterns at various times of the day (i.e., different time bands).

(i) Consumption seasonality

The customers' seasonal consumption patterns vary. They follow the typical Nordic phenomena: electricity consumption is relatively higher in cold winter months than in summer time. This can be seen from both sets of seasonal consumption variables (see Figure 12 a. & b.). However, it is important to note that there is a special group of customers in cluster IV (see Figure 13), whose electricity consumption in May-September is higher than that of the rest of the customers within cluster IV. This special group can be identified both from Figure 12a. (May-Sep. Consumption) and Figure 12b. (SummerConsumption), which emphasizes that the consumption deviation of this special group of customers in summer time is regardless of the summer months partition (i.e., May-September vs. March-September). At this point, it demonstrates that such a SOM-based data mining approach can visualize latent information for companies to take action upon. For instance, it could be of interest for the case company to investigate the phenomenon of this special customer group. However, the data we can access does not support to investigate further and provide certain explanation.



a. Conventional variable set b. Proposed variable set

Figure 12. Seasonal Consumption Visualization

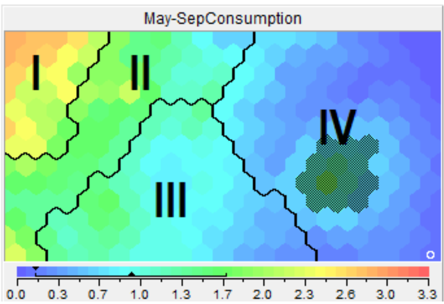


Figure 13. Special Customer Group Visualization

Based on the SOM visualization results, Figures 14 and 15 summarize the comparison of various time series profiles among clusters. Figure 14 illustrates the consumption profile breakdown of each cluster and the special group within cluster IV, regarding weekday/weekend as well as seasonal consumption patterns. The different clusters have distinct consumption profiles in different seasons. For instance, regarding the medium-low consumption customers (cluster III), their electricity usage is relatively even across different seasons (Jan-Apr., May-Sep. and Oct.-Dec.) in 2009 (red line in Figure 14). But high and medium-high consumption customers (purple and green lines in Figure 14) had lower electricity consumption in summer time, compared to their respective cold weather seasons. On the other hand, for low consumption customers, their consumption between May-September is relatively higher than in other seasons (see two blue lines in Figure 14). This is especially visible regarding the special group (light blue line in Figure 14) within this low consumption cluster.

Additionally, it needs to be noted that there is no significant variation between weekday and weekend consumptions within each cluster. This implies that in order to shift customers' demand between weekday and weekend to mitigate system constraints or when the wholesale market price is high, the utility should devise enough incentive in their price signals for customers to adjust their consumption behavior.

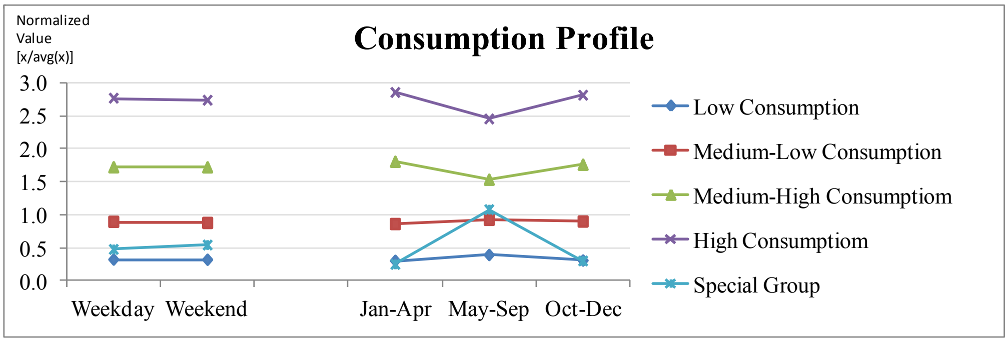


Figure 14. Consumption Profile Breakdown

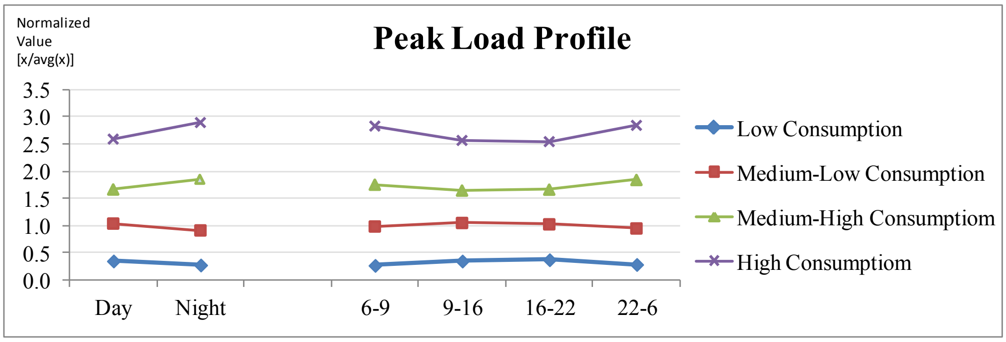


Figure 15. Peak Load Profile Breakdown

Note: the avg(x) in Figure 14 & 15 is calculated based on the respective seasonal/time period (e.g., Jan-Apr, May-Sep, Oct-Dec, weekday, weekend, etc.), as explained in Page 49.

(ii) Load patterns at various times of the day

Figure 15 shows a different picture regarding Peak Load at Day- and Night-time on the left hand side compared to that in four time bands on the right hand side. For instance, the peak load of high consumption customers (purple line on the left) is slightly higher at night (23:00-7:00) than during day time

(7:00-23:00). However, if one looks at the four time bands on the right, the purple line (high consumption customers) bends up towards the ends considerably. It indicates that the high consumption customers have relatively higher peak demand during the time periods 6:00-9:00 and 22:00-6:00, compared to usual working hours (9:00-16:00) or the usual peak consumption time period (16:00-22:00). The finding suggests that using the four time bands can reveal more detailed information about the customers' consumption behavior.

4.2 Implications

Even though the consumption profiling conducted in this research is not in 'real time' and is still based on historical data, it differs from load profiling in the traditional sense. By definition, a load profile is the '*estimated* use of energy by a customer or group of customers for each hour of the day' (Bailey, 2000). Consumption profiling reflects *actual* energy usage by customers, depending on the metering intervals.

In traditional practices, the electricity utilities have classified customers according to their business nature (i.e., industrial, commercial, and residential), the customer's consumption bands (e.g., annual consumption < 2,000 kWh, >5,000 kWh, or >18,000 kWh), and the housing types for household customers (e.g., detached houses, town houses, and multi-storeyed buildings). Even in the same customer class, the consumption patterns may vary considerably due to customers' business nature or life style diversity (Keppo and Räsänen, 1999). Additionally, the customer type is usually determined when the electricity connection is contracted, which is highly likely outdated because of later changes in the customer's profile, for example, occupancy changes in a household. Now, smart meter data provides the opportunity to group and compare the customers according to their actual energy usage, especially taking seasonal variations into account.

The empirical studies indicate that this SOM-based descriptive analytical approach is capable of visualizing deviations and more detailed information regarding customer's consumption patterns, which could support the utility in pricing differentiation. For instance, the result as reported in Paper II shows that there is a special group of customers within the low consumption cluster IV, whose consumption pattern in summer time deviated from that of the rest of the cluster. Based on this, the utility might design a particular product (e.g., solar energy), price (e.g., summer variable rates), or new product/price mix campaign to this group of customers.

Moreover, the consumption time series profiles of different customer clusters, as presented earlier, could be used as an indicator for designing various Time-of-Use (TOU) tariffs. In addition, there is evidence that the use of four time bands provides granular information regarding customer consumption behavior, which could aid the utility in designing demand response programs. Alternatively, it might be beneficial if the utility would consider using more than two time bands in the TOU pricing. These findings are actionable information for the utility to take into account in their future pricing decision-making.

On a related note, the comparison of electricity contract choice to the customer's actual consumption patterns, as in Paper I, reveals that the majority of customers with high consumption profiles chose the normal rate, instead of the economic rates or time rates which are designed in favor of high consumption customers. The reason behind might lie in the fixed components of the tariffs, but illustrates well how consumption profiling could help the utility to identify potential business opportunities, e.g., in terms of targeted retailing.

The empirical studies suggest that the SOM-based customer consumption behavior profiling method can provide added value to DSOs or electricity suppliers (retailers). Additionally, it implies that analyzing customers according to their consumption similarity and/or deviation could assist DSOs and electricity suppliers (retailers) to develop a better understanding of their customers, which in turn could not only aid them in dynamic pricing[6] which will facilitate demand response applications, but also allow them to better serve customers' consumption needs in order to secure the quality of supply.

Here both DSOs and electricity suppliers are mentioned because of a number of reasons. First, in many cases, a utility serves both distribution and retailing, even though functionally they are separate units. This is particularly true with small regions, such as the case company. Second, it concerns which entity owns or can access the smart meter data in order to perform the customer consumption profiling. Depending on different smart meter roll-out arrangement, it could be DSOs or electricity suppliers. Third, the role of DSOs is evolving, especially with the smart grids transition. The underlying assumption is that if the DSO wants to actively engage in various demand response applications, utilizing the smart meter data for consumption pattern profiling will allow understanding the customers' demand in near real-time.

[6] Dynamic pricing refers to pricing signals that are triggered based on actual wholesale market prices and not set in advance. For example, a Time-of-Use (TOU) rate is not a dynamic price, because the peak period rate and timing are set in advance. Critical Peak Pricing (CPP) is dynamic, because while the rate may be set in advance, the critical days are called based on wholesale market conditions (Faruqui and Wood, 2008)

Regarding the added value, for instance, under the common arrangement, DSOs owns the smart meter data, thus DSOs can form certain partnership with electricity retailers in demand response programs by providing the consumption profiling (new business opportunity for DSOs).

4.3 Chapter Summary

This chapter exemplified a descriptive analytical approach of BI, namely consumption profiling, based on visual clustering by using the SOM. A set of data mining models were built, one of which was detailed in the chapter, in order to demonstrate the added value which could be provided by BI approach. As such, this chapter addressed **RO1** and **SO1**. In the following chapter, an enterprise-wise real-time BI architecture will be presented.

Chapter 5

A Real-Time BI Scenario

This chapter presents an enterprise-wise real-time BI architecture based on Paper III. With the growing deployment of Advanced Metering Infrastructure (AMI), we believe that integrating smart meter data into legacy enterprise data warehouse systems and turning near real-time data into market intelligence for business innovation is crucial for the overall success of AMI investments. Taking a holistic view of BI, this chapter discusses how to turn smart meter data into pricing intelligence for innovative business development. It will begin with a brief overview of the BI concept in the IS field, and then describe the proposed real-time BI enterprise architecture.

5.1 Brief Overview of BI Research

In the decision support domain of IS research, business intelligence (BI) has evolved as an important field over the past decade. Meanwhile, in the world of practice, BI has been recognized as a strategic initiative and a key enabler for driving business effectiveness and innovations. BI, as defined in Chapter 1, is an umbrella term to describe methodologies, architectures, applications, technologies, and processes for gathering, storing, accessing, and analyzing data to help business users make informed decisions (Wixom and Watson, 2010). BI involves database management system (DBMS) and data warehousing, usually classified as 'getting data in' on the one hand; and enterprise reporting, online analytical processing (OLAP), querying, business performance management (BPM), data mining, complex event processing (CEP), etc., as 'getting data out' on the other (Watson and Wixom, 2007). It is widely acknowledged that 'getting data in' delivers limited value to an organization (Chaudhuri et al., 2011; Watson and Wixom, 2007), in terms of performance improvement, decision making support, and innovation creation. In other words, the purpose of BI is not only concerning establishing IT-enabled enterprise infrastructure; it is also aiming to facilitate business innovation and to transform an organization towards sustainable profitable growth.

As a research subject, BI has been extensively studied in the IS discipline. For instance, Jourdan et al. (2008) analyzed 167 articles with BI related topics in ten leading IS journals published from 1997-2006. During the ten-year period, five distinct BI research categories and nine BI research methodologies can be identified within IS research. This speaks of the growing scale, scope, and importance of BI research.

Recently, Maghrabi et al. (2011), Saldanha and Krishnan (2011), and Marjanovic and Roose (2011) examined the strategic and operational role of BI in service innovation, product and service development, and business process improvement, respectively. Watson (2009), Wixom and Watson (2010) proposed the concept of BI-based organizations because of the increasing role of BI in the enterprise's operations and overall business success, as seen in Harrah's Entertainment and Continental Airlines. Baars and Zimmer (2013) and Zimmer et al. (2012) have elaborated the inter-relation between the effectiveness of BI and agility measures. The above-mentioned studies have shown BI capabilities of supporting problem and opportunity identification, decision making, and alignment of operations with corporate strategy, thus contributing to the enterprise's competitiveness and sustainable development (March and Hevner, 2007; Olbric et al., 2011).

5.2 Real-Time BI Enterprise Architecture

Throughout history we have seen many successful stories regarding how technology advances can facilitate new business models. The advent of the Internet and the mobile revolution in Africa are two good examples. How can the advance in smart metering technologies be transformed into business innovations in electricity supply business? A potential answer is real-time[7] BI enterprise architecture.

As shown in the previous chapter, smart metering technologies have enabled timely customer consumption behavior analysis in the form of visual data mining. Such a descriptive analytics based BI approach could guide the management team to formulate new pricing strategies. According to the findings of consumption profiling in Chapter 4, for instance, the management team could consider whether it will be beneficial to offer a special electricity tariff to customers who have higher summer consumption needs. Or, to which extent does the price signal provide adequate incentive to steer consumption shifting from peak period to off-peak period in dynamic pricing?

In order to answer the questions raised above, a holistic BI approach is required. As illustrated in Figure 16, in addition to the AMR and customer

[7] Some people argue that data for real-time BI only need to be as fresh as the decision or business process require. Depending on the business need, data can be hourly, daily, and even weekly or monthly and still be real-time (Anderson-Lehman et al., 2004).

demographics systems, the operational systems that contain other internal data such as financial data, service data, and billing system data need to be included. Moreover, external data sources such as regulatory data, weather data, wholesale market price, etc., also need to be integrated through data warehousing, in order to support designing new product/price mixes, or innovative business models. In particular, the Complex Event Processing (CEP) engines, in-memory databases, and column stores will play an important role in supporting near real-time decision-making, as explained in the following.

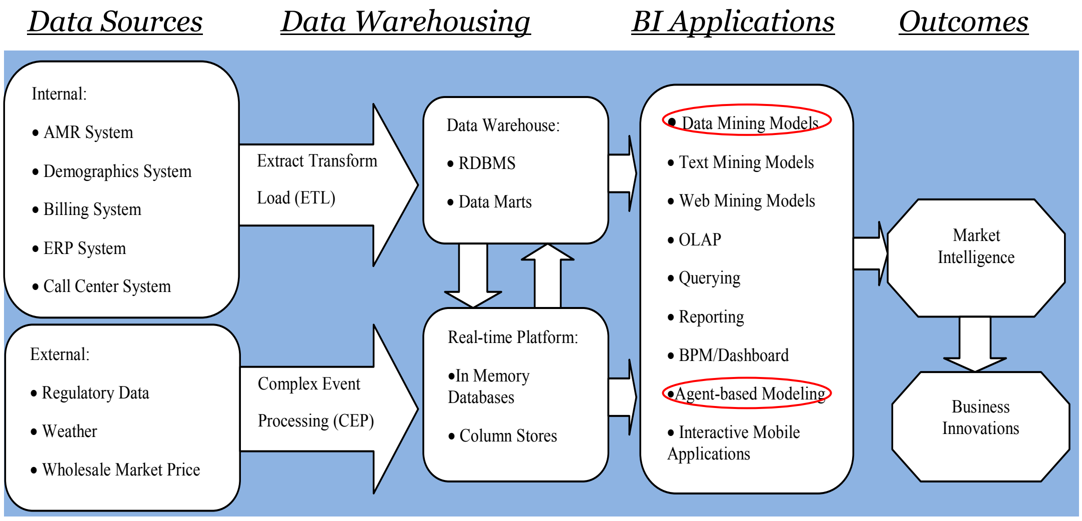


Figure 16. Real-time BI Framework

With respect to near real-time decision-making, reducing the time taken to collect and store the data (i.e., *data latency*), as well as the time taken to analyze the data and turn it into actionable information (i.e., *analysis latency*), is prerequisite (Hachathorn, 2004). Traditional data warehousing architectures deploy a DBMS, in which disparate data sources are integrated through the data extraction, transformation, and loading processes (ETL) into time-variant repositories (such as data marts). The increasing volume of streaming data (such as from smart meters), and the velocity of streaming data with sudden variations in data processing (such as real-time pricing), have challenged the conventional data warehousing approach in terms of computational efficiency in data storage, data retrieval, and querying.

CEP, in-memory databases, and column stores belong to an emerging set of concepts and techniques in data warehousing, aiming to address the needs for large volume data storage and real-time data access and analysis. For instance, *column store*, also known as column-oriented database system, is a database management system (DBMS) that stores data tables as sections of columns of data in comparison to traditional row-oriented database systems. Column stores are particularly suitable for read-mostly, read-intensive, large data repositories (Abadi et al., 2009). Unlike traditional DBMS that typically employ a disk storage mechanism, *in-memory database* is a DBMS that use a system's main memory for data storage. Essentially, in-memory databases facilitate efficient data access and database requests in high volume environments where response time is critical.

CEP, on the other hand, differs from traditional data warehousing since operational data does not need to be first transformed and loaded into a database before it can be analyzed (Chaudhuri et al., 2011). *CEP* is a new type of computing for monitoring and processing streams of data from multiple sources so as to infer patterns or events in real-time (Luckham, 2012). The goal of CEP is to (1) identify a 'change of state' (i.e., event) when a measurement exceeds a predefined threshold, (2) detect complex events by analyzing and correlating other events, and (3) extract information from events as they occur that can be utilized to take immediate actions in response to those events (Luckham, 2002). CEP relies on a number of techniques[8], including (Etzion and Niblett, 2010):

- event-pattern detection;
- event abstraction;
- event filtering;
- event aggregation and transformation;
- modeling event hierarchies;
- detecting relationships (such as causality, membership or timing) between events;
- abstracting event-driven processes.

As explained above, CEP engines, in-memory databases, and column stores are important components in the real-time BI framework contributing to minimizing *data latency* and *analysis latency*. These new techniques in data warehousing make real-time analytics and BI applications possible. For instance, if the real-time BI architecture is in place, and provided with hourly AMR data, it will allow the company to update their understanding of the customer's consumption profile in a near real-time manner. With the gained better knowledge of customers' electricity demand, the company's decision makers can differentiate their pricing strategy when such a business need arises. For example, in the demand response scenario, the company can suggest a lower tariff at a certain time of the day to a particular customer group, in order to serve the purpose of mitigating electricity demand from a critical peak period to off-peak time. The selection of the particular customer group is based on up-to-date and accumulated knowledge of the customer's consumption profile and the willingness and feasibility of the customer to

[8] For thorough discussion of these techniques, please refer to Etzion and Niblett (2010).

respond to the price signal. The company can specify the event patterns and certain profiles in the rule sets. When the specified patterns are identified over a high-volume stream of events by the CEP engine (e.g., extreme weather or wholesale market price surge), it will trigger appropriate responses.

Such a capability of steering electricity demand is crucial for the electricity utilities to perform under the complex demand-supply dynamics of electricity supply. It is also significantly important with respect to integrating the intermittent renewable energy sources into the electricity supply scene, which is another business challenge faced by the electricity industry. The real-time BI enterprise architecture is imperative in that it can enhance the management capability to achieve the goal 'from intelligence to innovation' for sustainable growth. In other words, the real-time BI enterprise architecture supports in turning near real-time smart meter data into market intelligence, thus facilitating business innovation in the electricity retail market.

It is important to note that in terms of real-time BI there are three types of latency (i.e., *data latency*, *analysis latency*, and *decision latency*), as described in Chapter 1. *Data latency* and *analysis latency* could be tackled by real-time BI enterprise architecture, employing new data warehousing technologies, and advanced analytics and BI applications. However, the real-time BI enterprise architecture can only assist in reducing *decision latency* to certain extent, i.e., providing actionable information to support decision making in a timely manner. Whether and how the business decision makers will react to such information in their decision-making process is not what the real-time BI enterprise architecture can affect. This belongs to the domain of organizational studies.

This doctoral study is mainly focused on two types of analytics for real-time BI applications, namely data-mining based descriptive analytics and agent-based prescriptive analytics. The descriptive analytics based on visual data mining, as presented in the last chapter, provided empirical support for deriving the real-time BI framework in the current chapter. Prescriptive analytics will be presented in the next chapter via agent-based modeling and simulation. On the one hand, agent-based prescriptive analytics demonstrates how flattening daily load patterns can be achieved by customers adjusting their consumption, given that the real-time BI enterprise architecture is in place. On the other hand, it serves as a proof-of-concept in terms of the proposed real-time BI framework. Implementing and evaluating the real-time BI framework requires further investigation, especially through field testing. This could be a candidate for future research.

5.3 Chapter Summary

This chapter presented a conceptual framework of real-time BI enterprise architecture. This is based on the understanding that real-time BI has been recognized as a critical enabler of value creation and business innovation both in the IS literature and in practice. We argue that holistic real-time BI enterprise architecture is the key for innovative initiatives, particularly relevant to dynamic pricing. And the BI-based management capability will lead an organization to sustainable profitable growth in the transition towards a demand response electricity retail market. This chapter contributed to addressing **RO2** and **SO3**. In the next chapter, an agent-based price-responsive demand simulation will be explained, in order to demonstrate a prescriptive analytical approach of BI.

Chapter 6

Agent-Based Modeling for Price-Responsive Demand Simulation

This chapter describes an approach for prescriptive analytics through agent-based simulation. It is based on Paper IV. In terms of the whole research design, the agent-based modeling and simulation serves the purpose of justifying the real-time BI framework presented in the previous chapter. This chapter is organized as follows: first, the experiment setup will be briefly introduced; then, the multi-agent-based meta-model (MAMM), the domain model (DM), and a use case will be presented; and finally, the results and implications will be discussed.

6.1 Experiment Setup

As presented earlier, the ongoing deployment of Automated Metering Infrastructure (AMI) not only opens up the opportunity for electricity consumption profiling, but also creates a platform for dynamic pricing and demand response applications. In order for the demand responsive paradigm to be realized, it is crucial to understand the ever-evolving interactions between the demand and the supply sides in the electricity retail market. The competitive electricity retail markets are relatively young and the restructured electricity markets are inherently complex – encompassing physical constraints, different institutional arrangements of operations, and varying behavioral dispositions of market participants. There is an increasing need for advanced modeling approaches that simulate the emergent behavior (demand responsiveness) among market participants (e.g., consumers, suppliers, producers, etc.).

Agent-based modeling (ABM), compared to traditional system-modeling techniques, is one promising approach for studying how the market participants might act and react to the complex economic, financial, regulatory, and environmental circumstances embedded in the electricity sector. ABM has been increasingly studied for the simulation of electricity markets in recent decades along with the electricity industry restructuring and unbundling. Very often the demand side is represented as a fixed and price-insensitive load (Weidlich and Veit, 2008). In order to develop dynamic pricing solutions, it is important to take the varying cost of electricity in the wholesale market, as well as the level of demand participation, into account, especially with regard to household customers and small and medium-sized businesses.

One of the unknowns in implementing dynamic pricing is whether and by how much customers would reduce peak loads in response to changing price signals (Faruqui and Wood, 2008). Therefore, it would be necessary to estimate the impacts of time-varying electricity rate on peak-period energy use, before any experimental pilots can be carried out. Quite often, it is complicated, costly, and time-consuming to organize this kind of pilot pricing program. Agent-based modeling and simulation will allow observing and growing the interactions between pricing and demand in a controlled experimental setting.

Therefore, in this research, a multi-agent-based modeling and simulation experiment was constructed for systematically modeling price-responsive emergent behavior in the context of the demand response electricity retail market. Firstly, a multi-agent-based meta-model (MAMM), which defines the concepts, relations, and structure of agents on an abstract level being independent of any concrete domain, was introduced. Secondly, the MAMM was instantiated with domain specific notions, which provide a uniform abstract interpretation of all domain models that conform to the MAMM. Thirdly, a specific simulation experiment was formulated accordingly. Taking a prescriptive analytical approach, the current experiment demonstrated how the MAMM guided domain model construction can be used to address the pricing impacts on demand by means of simulation.

6.2 Multi-Agent Based Meta-Model (MAMM)

The MAMM is a customized version of *utility-based agents meta-model* introduced by Russell and Norvig (2003). The underlying assumption is that the decision-making agents have *bounded rationality*, and their decision making is according to a *utility function* which maps a state to a desirable situation. The MAMM is illustrated in Figure 17 as a semantic network, and is depicted as follows.

An *agent* has one or more *roles*; each of these roles determines one or more *goals*. The way how an agent reacts to the environment/other agents with different actions depends on the mode and its goal. A *mode* includes a set of agent's states. To fulfill its role an agent performs *actions* that are *triggered* by some *event*. The actions, in turn, can *generate* new events when terminating (atomic actions) or in the course of execution (non-atomic actions). An event is a notion related to both – *time* and *state*. An event reflects the instant of time when some change of state occurs. A state is defined as a valuation of agent attributes. State is changed by actions. Action may have non-zero extent in time. Since each action describes only a subset of *state changes*, the action is *enabled* only in certain states. A set of actions is called *interaction* if the agents' actions on shared states are in the '*changes*' and '*depends*' relations.

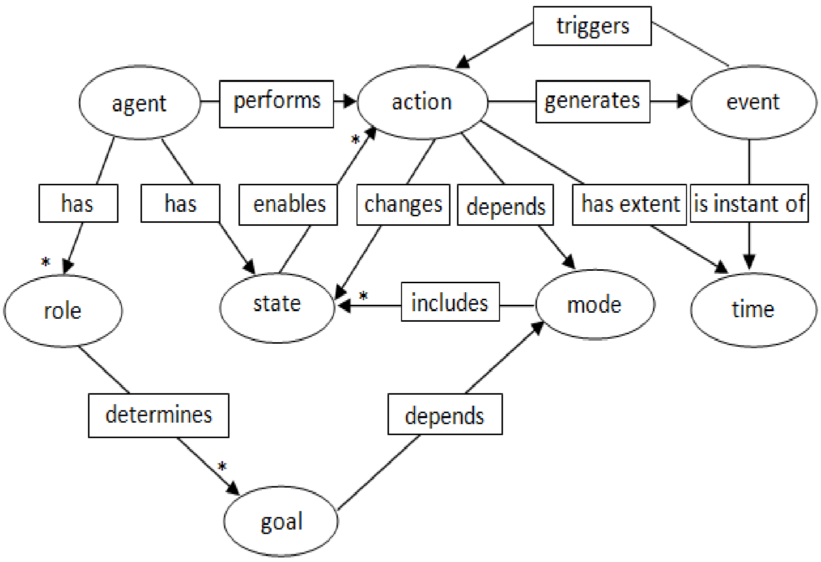


Figure 17. Semantic Network of the MAMM

6.3 Domain Model Instantiation

Each domain model (DM) that conforms to the MAMM is considered as an instantiation of the MAMM. Since the domain model usually includes multiple notions all being instances of meta-notions of the MAMM, the concrete selection of the domain concepts depends on the specific analysis problem for which the domain model has to be constructed. The following DM instantiation is for price-responsive demand analysis.

The agent is to represent the market participants in the real world and act on behalf of them. In the context of electricity markets, it includes producers, transmission and distribution operators (TSOs and DSOs), electricity suppliers (retailers), consumers, and other load serving entities (e.g., demand aggregators). Even though the environment is external and largely uncontrollable, it must be simulated as an agent in order to show how it will affect production and consumption activities of the market participants.

For price-responsive demand modeling, a domain instantiation can be characterized as in Figure 18 a, b, & c. Since the consumer and the supplier are the focal market participants in this context, the focus of the DM is on their actions, states, and interactions.

Actions

The supplier's major business activities (see Figure 18 a.) include (1) pricing in the retail market (i.e., offering various retail electricity rates to different consumer groups) according to the supplier's market share and profit maximization objectives; (2) bidding in the wholesale market, which will generate the following day's hourly spot price; and (3) hedging in the financial market in order to avoid the risk caused by energy price volatility.

The consumer's activities in relation to electricity consumption (see Figure 18 b.) include (1) consuming electricity according to their business nature and living needs; (2) analyzing the possible saving from choosing the demand response tariff, and the feasibility and the cost/inconvenience of rescheduling electricity consuming activities in order to respond to changing price signals (i.e., cost-benefit analyzing when facing time-varying price or demand response tariff); (3) adjusting timing and level of consumption based on real-time information and price signals.

States

The supplier's initial pricing action is determined by their state. Various ownership relations, different marketing and risk management strategies, the supplier's market share and profit maximization objectives, and the supplier's electricity rate portfolio and dynamic pricing program design will determine the supplier's state (see Figure 18 a.). The varying state, in turn, will have an influence on the supplier's pricing practice.

Similarly, the consumer's state (see Figure 18 b.) will determine the consumer's actions in terms of electricity consumption and the possibility to respond to dynamic pricing. The varying demographic attributes (e.g., price sensitivity, risk preferences, and the composition of electric appliances), the feasibility to shift certain electricity usage to off-peak time, the perceived saving, the rescheduling cost, the tolerance towards inconvenience, and so on, will all affect the consumer's price responsiveness when the consumer is facing a new pricing offer.

Interactions

As shown in Figure 18 c., the supplier's pricing based on the wholesale market price (action 1) will generate a new retail price (event 1). It will trigger the consumer's action of analyzing cost-benefits (action 2). Here an interaction occurs in response to the supplier's pricing activity (interaction 1). After analyzing the new price signal, if it is not feasible to shift the demand, the consumer will remain the original state and keep the electricity consumption as usual (action 3). If it is feasible to shift the demand, it will lead to the consumer's state change, which in turn will enable the consumer's action of adjusting consumption (action 4).

The adjusted electricity consumption is the consumer's price-responsive demand (event 2), which will have impact on the supplier's bidding activities for the following day (action 5). Accordingly, the new spot price (event 3) resulting from the current interaction (interaction 2) will trigger the next round interaction between the supplier's pricing activity and the consumer's cost-benefit analyzing and electricity usage adjusting (if possible) activities.

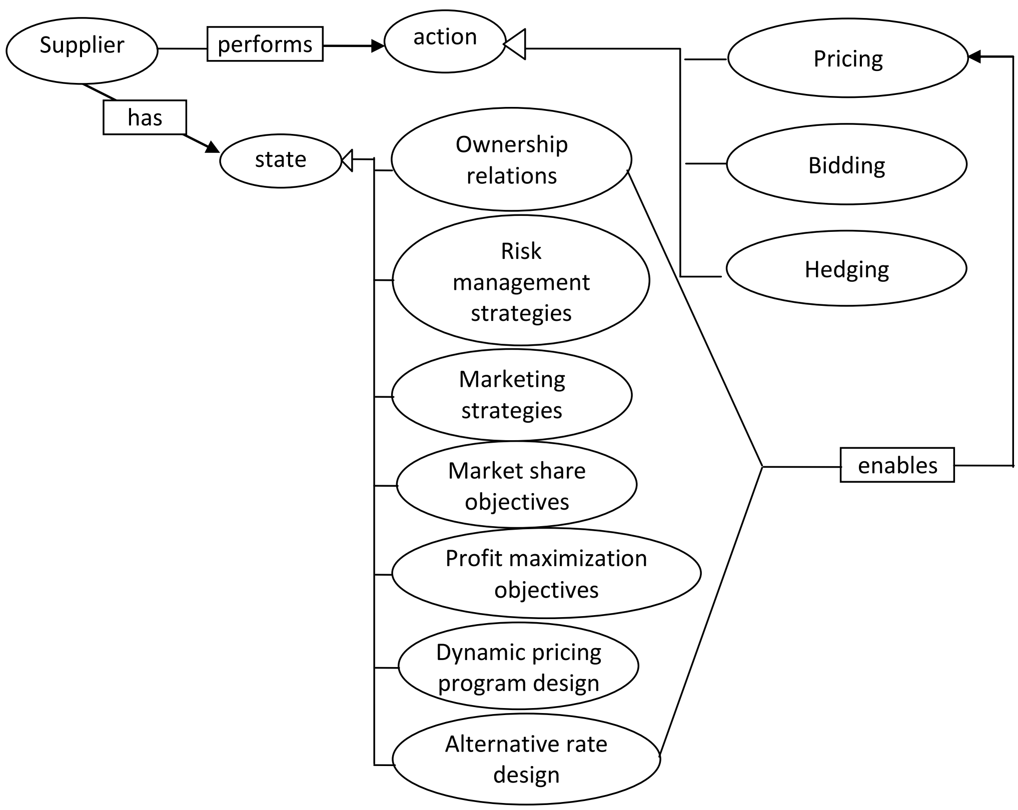


Figure 18a. Supplier's Action-State Diagram

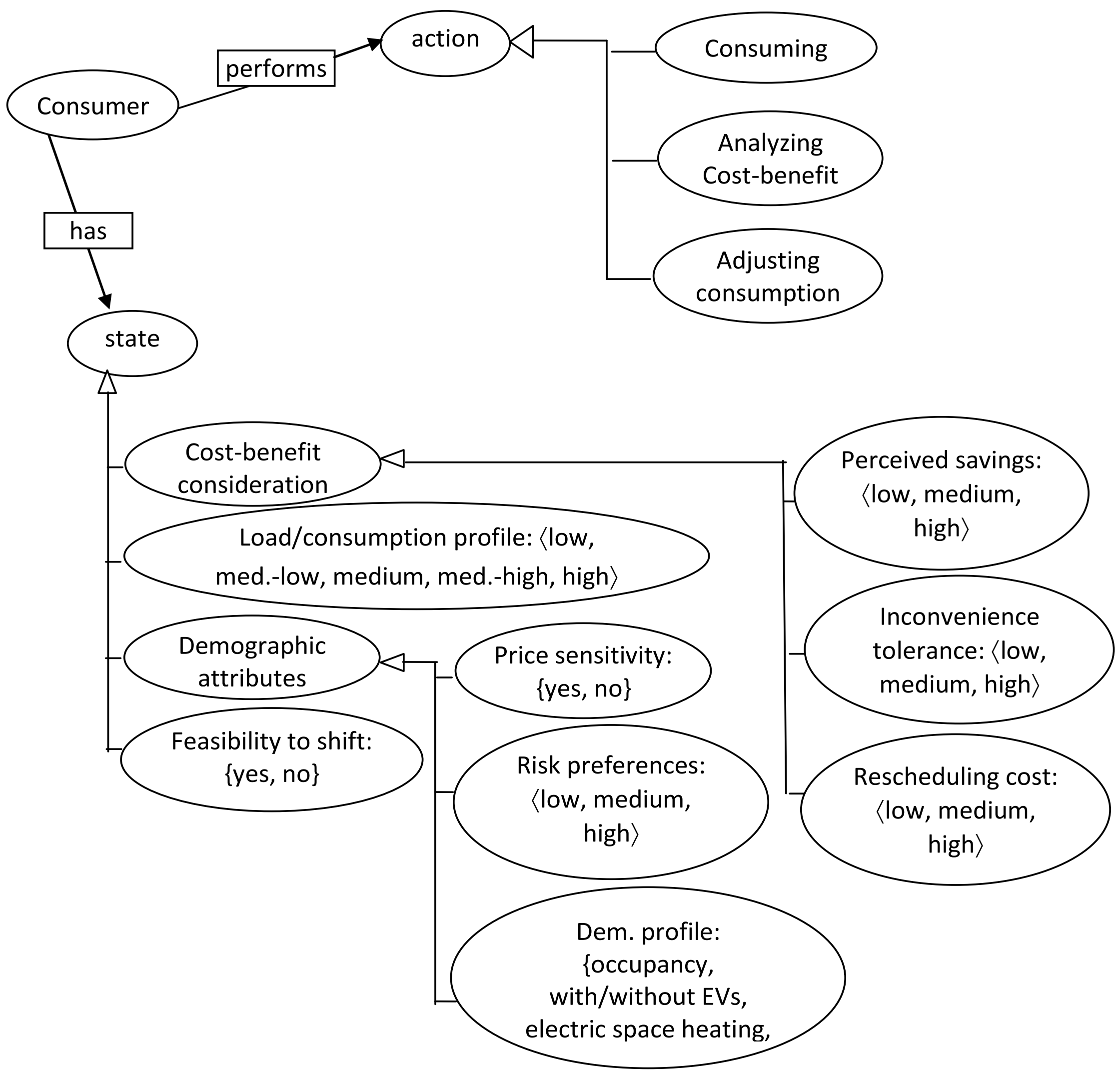


Figure 18b. Consumer's Action-State Diagram

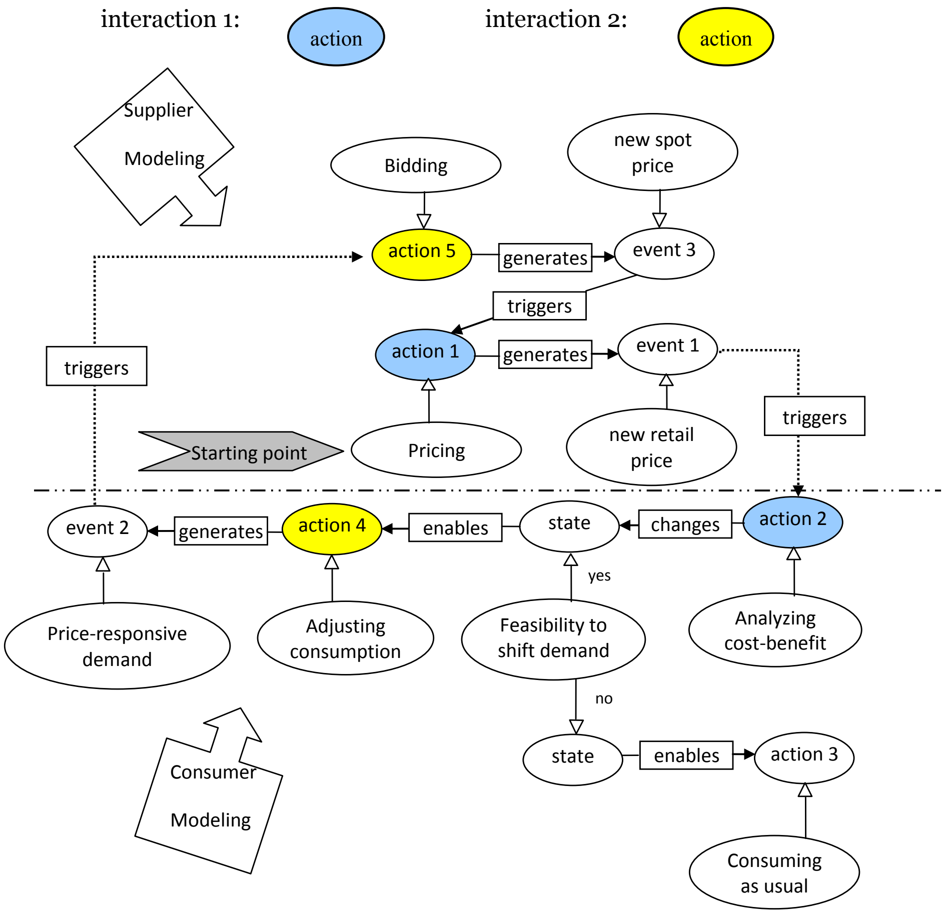


Figure 18c. Supplier-Consumer Interaction Diagram

6.4 Use Case

Based on the domain model described above, simulation experiments can be carried out. The specific theoretical simulation scenario is for the supplier to obtain the ideal demand curve which has such a shape that it follows the spot price curve in inverse ratio (Belonogova et al., 2011). The simulation model represents the Supplier-Consumer interaction as in Figure 18c., where the interaction observables are hourly price and hourly consumption. The simulation model is formalized and run in the UPPAAL environment (Bengtsson and Yi, 2004), which is an academic-free modeling, simulation, and model-checking tool.

The simulation setup consists of 1 supplier and N consumers. The consumers belong to a high consumption cluster (HCC), which makes steering their demand according to the spot price a priority in relation to the supplier's goal of profit maximization. The spot price is based on the Nord Pool Spot published system price for Estonia during the 2nd week of January, 2013 (www.nordpoolspot.com).

Pricing Function

The pricing function is designed with the following assumptions. The supplier calculates hourly price based on spot price in order to smooth sharp fluctuations in consumption without altering the HCC's total consumption and possibly increasing the supplier's profit. Also, an upper limit Δ^{TL} to hourly price change Δ is set in order to avoid overshoots and instability of consumption. The HCC consumers adjust their daily electricity consumption activities (if possible) in response to the price signals.

The basis of the next day's hourly price *P'(T)* at hour *T* is the spot price *P (T)* of the previous day at *T*. Let *Q (T)* be the consumption at *T* on previous day. Then the next day's hourly price *P'(T)* at hour *T* is calculated by formula (6.1).

$$P'(T) = P(T)\,(1 + \Delta(T)/100), \text{ where} \tag{6.1}$$

$$\Delta(T) = \begin{cases} \dfrac{v \cdot [P(T)\cdot Q(T) - avg(P(T)\cdot Q(T))]}{avg(P(T)\cdot Q(T))}, & \text{if } \dfrac{\Delta(T)}{P(T)} < \Delta^{TL} \\ sign(\Delta(T)) \cdot \Delta^{TL} \cdot \dfrac{P(T)}{100}, & \text{otherwise} \end{cases} \tag{6.2}$$

where:
v is parameter to amplify or suppress the effect of calculated price correction;
Δ^{TL} is acceptable price change (%);
$sign(\Delta(T))$ is the sign function with co-domain {-1, 1}, showing if the price correction is positive or negative comparing to previous day's spot price.

The hourly price calculated by using formula (6.1) is proportional to the difference $P(T)\cdot Q(T) - avg(P(T)\cdot Q(T))$, where $avg(P(T)\cdot Q(T))$ is the arithmetic mean of $P(T)\cdot Q(T)$ over 24 hours. Formula (6.2) guarantees that the calculated change of hourly price never exceeds the limit set by Δ^{TL}. This is needed for maintaining the stability of price response.

Consumer's Behavior

The consumption patterns in the HCC include consumption activities, e.g., ironing, room heating, water heating, etc. Each activity is characterized by the following attributes: enabling conditions and consumption interval or function, as summarized in Table 3.

Table 3. Descriptive Attributes of the HCC's Consumption

Activities	Enabling condition(s)			Consumption interval/ func. (Wh)
	Time interval	Price zone	Outdoor temp.	
Laundry, dish-washing	00 - 24	$P \in Z_1$	-	$[C_1,C_2]$
Ironing	19 - 22	$P \in Z_1 \cup Z_2$	-	$[C_3,C_4]$
Water heating	06 - 23	$P \in Z_1 \cup Z_2$	-	$[C_4,C_5]$
Cooking	07 - 08; 18 - 19	$P \in \cup_{i=1,5} Z_i$	-	$[C_6,C_7]$
Lighting	07 - 09; 18 - 24	$P \in \cup_{i=1,5} Z_i$	-	$[C_8,C_9]$
Space heating	00 - 24	$P \in \cup_{i=1,3} Z_i^c$	$T < T_{crit}^a$	$E^b (T_{crit}-T)$

Note:

a. T_{crit} is the highest outdoor temperature when the space heating is activated (e.g., T_{crit} = 16 °C).

b. E is the amount of energy needed for space heating in order to compensate the decrease of outdoor temperature by one degree (e.g., E = 50 W/ ° C).

c. Z_i represents price zone, where Z_1 = [<, 34], Z_2 = [35, 39], Z_3 = [40, 44], Z_4 = [45, 49], Z_5 = [50, >] (EUR/MWh).

In the simulation model, when consumption dependency is well-defined it is specified by means of explicit function such as in space heating. Otherwise, the consumption is approximated by conditional probabilistic functions where each function returns a random value from the consumption interval, e.g. [C1, C2] for laundry and/or dishwashing, [C8, C9] for lighting. The distribution within the consumption intervals is assumed to be discrete uniform distribution because of the UPPAAL environment. The consumption intervals are aggregated based on typical energy usage of household appliances. For instance, the consumption interval for lighting [C8, C9] ranges from 500-650 Wh.

6.5 Results and Implications

Simulation Results

The simulation indicates that in the presence of HCC consumption patterns the implemented pricing strategy allows smoothing the demand peak in relation to the spot price.

Figure 19 shows the dynamics of pricing-demand interplay: P is the curve of the spot price of Jan. 09, 2013, and P' represents the hourly price curve generated by the model as described in formula (6.1). If the price is decreased from 44 to 42 EUR/MWh at off-peak time period (11-17hrs), it will encourage considerable demand shifting to this period (from 500 to 1000 MWh). On the contrary, if the price is increased during the spikes of Q from 40 to 44 EUR/MWh at 19hrs and from 34 to 38 EUR/MWh at 22hrs, it will cut the demand to Q' (from 2600 and 2800 to 2200 MWh).

The specified pricing strategy demonstrates the effect of flattening the daily load. The standard deviation of the demand Q' decreases about 57 % in comparison to demand Q.

It is important to note that the simulation is based on a theoretical scenario. It does not take into account the impact of other market participants' activities such as the producer's actions and other environmental factors except the outdoor temperature caused spot price change and demand adjustment. In addition, the agent capacity of learning and adaptation is not considered in the simulation due to short time range.

The UPPAAL environment includes a verification tool, called 'verifier', which is 'to check the safety and liveness properties by on-the-fly exploration of the state-space of a system'. Therefore, the model-checking is performed within the UPPAAL environment.

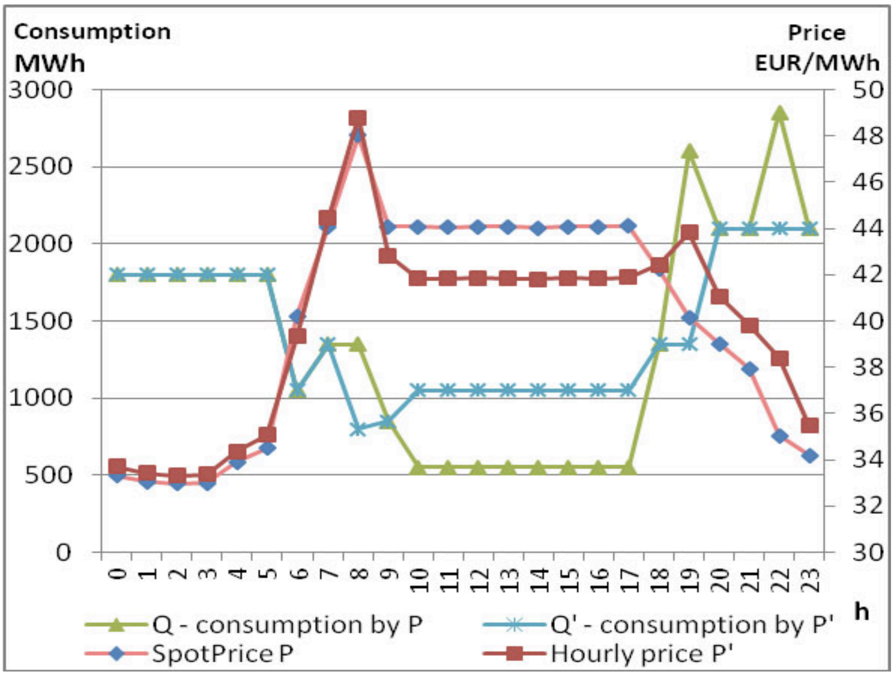


Figure 19. Price-Responsive Demand

Implications

The MAMM guided domain model instantiation provides a conceptual platform to model the electricity retail market evolving interactions in a systematic manner. Based on the domain model and its formalized representation, simulation experiments can be developed. The use case shows that the DM is rich enough in order to validate the conceptual construct and these constructs provide a set of model patterns that are easy to handle when formalizing the domain model.

By definition, prescriptive analytics is to recommend decision options and to show the impact of each decision option using optimization, simulation, etc. The ABM based approach provides an option to experimenting with various dynamic pricing strategies and examining the impacts on demand, and vice versa. For instance, in retail pricing practice in Nordic countries, with a *spot contract*, consumers will be charged the Nord Pool area price plus a margin or some other kind of commission. Consumers without hourly meter will be charged the unweighted monthly average of the system price (Johnsen and Olsen, 2011). The ABM-based pricing experiment could allow the electricity suppliers (retailers) to simulate various scenarios of margin vs. demand change in relation to spot price. By so doing, the electricity suppliers (retailers) could steer the demand with pertinent price signals reflecting the wholesale market situation and the company's profit maximization strategy.

This ABM experiment is formulated in the simplest case. The underlying consumption is that both the supplier and the consumer have *bounded rationality*. The supplier aims to steer the daily demand curve with incentive pricing signals, without compromising the company's profitability. The consumers are constantly attentive to the new pricing offer, and are willing to adjust their electricity usage if possible. As stated earlier, other factors that could affect the market dynamics are not included in the experiment design for the sake of simplicity. Since the primary purpose of the current study is to demonstrate an ABM based approach of prescriptive analytics in the context of innovative pricing.

6.6 Chapter Summary

This chapter demonstrated through agent-based modeling and simulation how daily load patterns can be flattened by consumers adjusting their electricity usage according to the spot market price. The ABM makes it possible to capture and observe the emergent behavior in the electricity demand and supply interactions. This aids to discover dynamic pricing solutions that reflect the varying cost of electricity in the wholesale market as well as the level of demand participation before any real-world costly pilot project can be implemented. Taking a prescriptive analytical approach, the agent-based price-responsive demand simulation to an extent served as a proof-of-concept in terms of the overall research design. This chapter addressed **RO2** and **SO2**, also contributed to **SO3**. The next chapter will present a SOM-based BI application for efficiency performance visualization regarding the electricity distribution business.

Chapter 7

Understanding the Regulated Distribution Sector

For the smart grids paradigm change to happen, the distribution system operators (DSOs) will play an important role. Therefore, it is necessary to include the distribution sector – which is a component part of the electricity retail market – in this study. As stated in Chapter 3, the local DSOs' operational efficiency determines the network tariff, which contributes to the retail price. This chapter presents a SOM-based application for visually benchmarking DSOs' efficiency performance, based on Paper V. It starts with the introduction of the background regarding electricity distribution benchmarking and the Finnish practice, then comes to the SOM model, results and analyses, and implications.

7.1 Regulatory Background

> "Electricity prices have risen by more than 50 per cent in real terms over the past five years. Spiralling network costs are the main contributor to these increases, partly driven by inefficiencies in the industry and flaws in the regulatory environment."
>
> "At this stage, benchmarking — which compares the relative performance of businesses — is too unreliable to set regulated revenue allowances. Nevertheless, greater and more effective use of benchmarking could better inform the regulator's decisions."
>
> -- Australian Government Productivity Commission: Electricity Network Regulatory Framework, April 2013.

As stated in Chapter 3, the purpose of electricity market deregulation is to introduce competition and market mechanism into electricity generation and supply, whereas the electricity transmission and distribution sectors are subject to regulation due to their monopolistic nature. Efficiency benchmarking of the distribution system operators (DSOs) has been adopted for economic regulation worldwide. The aim is to encourage cost-cutting and enable efficiency improvement, in turn to secure electricity supply at a reasonable price. At the same time, it provides indicators for energy market authorities to assess the regulatory policy such as the revenue/price cap or the return rate. The Australian Government Productivity Commission's report reflects the importance of benchmarking in this regards, and the interrelations between DSOs' efficiency, retail price, and regulatory decisions.

There are different incentive regulatory schemes, such as price/revenue cap, rate of return, yardstick, and so on (Jamasb and Pollitt, 2000; Joskow and Schmalensee, 1986). Rate of return (ROR) is a traditional cost-based regulatory scheme. It is also called cost of service regulation (Hill 1995). ROR allows the pricing of regulated utility to cover its operating and capital costs, plus a reasonable return on investment. By contrast, the price/revenue cap regulation is a performance-based scheme, the essence of which is to decouple the profits of the regulated firm from its costs by setting the maximum allowable price/revenue the utility can charge/earn. As a matter of fact, the distinction between cost- and performance- based approaches is the degree to which the costs are reflected in the network tariffs (Hill 1995). In yardstick regulation, the regulator establishes a 'performance benchmark'. For instance, the average costs of a group of comparable utilities can be used as a benchmark.[9]

Accordingly, the benchmarking methods used by respective national regulation authorities vary, including Data Envelopment Analysis (DEA), Corrected Ordinary Least Square method (COLS), Stochastic Frontier Analysis (SFA), and the so-called 'Network Performance Assessment Model' (Farsi et al., 2007; Honkapuro et al., 2004; Jamasb and Pollitt, 2001; Kinnunen, 2005). The following will concentrate on presenting the DEA method since it is relevant to the research topic at hand (i.e., the distribution regulation in Finland).

The Finnish practice in electricity distribution regulation and efficiency benchmarking has drawn special attention because the Finnish Energy Market Authority (EMA) has implemented a controversial regulatory scheme – the rate of return (ROR) scheme – which is deemed to lack incentives for cost-cutting (Hill, 1995; Jamasb and Pollitt, 2000; Kinnunen, 2005). Arguably, however, Finland has still achieved a more efficient electricity distribution system relative to many other European countries (Edvardsen and Førsund, 2003; Kinnunen, 2006). The Finnish energy authority has adopted the DEA benchmarking method with the ROR scheme since 2000. The direct outcome of the DEA-model is a score on a scale between 0 and 1 for each DSO in a given year, with the score of 1 denoting the most efficient DSO. Nonetheless, the scores only indicate which DSO performed efficiently or inefficiently. However, it lacks comprehensible connections to the actual operating

[9] See Joskow and Schmalensee (1986), Comnes et al. (1995), Hill (1995), Hall (2000), and Jamasb and Pollitt (2000) for further discussions about different incentive regulation models.

characteristics of the DSOs, in terms of the selected measurements (i.e., Operating Costs, Distributed Energy, Interruption Time, Network Length, and Number of Users). For instance, while the fully efficient firms form the efficiency frontier, there can be several different input/output combinations (i.e., operating circumstances differ) reaching full efficiency (see points A, B, C, D in Figure 20).

Data Envelopment Analysis

Data Envelopment Analysis (DEA) is a non-parametric and frontier-based method for quantitative performance evaluation. It has been widely applied in management science, economics, and operations research for the estimation of production efficiency. The principle of the DEA method lies in using piecewise linear programming to calculate the best input-output ratio (the efficiency frontier) in a multiple input and output case. Then the DEA benchmarks the relative efficiency of the units within an organization or the efficiency across firms against the best practice (i.e., the frontier). A mathematical formulation is available in (7.1), which is originally presented by Charnes et al. (1978).

$$
\begin{aligned}
&\text{Min } \theta && (7.1)\\
\text{s.t.}\quad & \theta x_{jo} - \textstyle\sum_{i=1}^{n} \lambda i x_{ji} \geq 0 \quad (j = 1, \ldots, p),\\
& \textstyle\sum_{i=1}^{n} \lambda i y_{ki} \geq y_{ko} \quad (k = 1, \ldots, r),\\
& \lambda i \geq 0 \quad (i = 1, \ldots, n).
\end{aligned}
$$

where :

θ: efficiency score for the assessed unit,

y_{ki}: observed output k of unit i,

x_{ji}: observed input j of unit i,

λ_i: weight of unit i in the reference point for the assessed unit,

subscript o: refers to the assessed unit, but the unit also preserves its original index.

On the one hand, the best practice frontier can be achieved through maximizing output for a given amount of input factors, the so-called output-oriented model. On the other hand, it can be obtained via minimizing the input factors required for a given level of output, accordingly, called the input-oriented model. Generally, the input-oriented model is used extensively in electricity distribution benchmarking (Jamasb and Pollitt, 2001). With respect to electricity distribution benchmarking, the efficient firms will form a frontier (i.e., the solid line A-B-C-D), which envelops the inefficient firms (e.g., E and F), as shown in Figure 20. Each firm's efficiency is represented by an efficiency score on a scale of 0 to 1, with the most efficient firms receiving the score of 1 (Jamasb and Pollitt, 2003; Korhonen and Syrjänen, 2003).

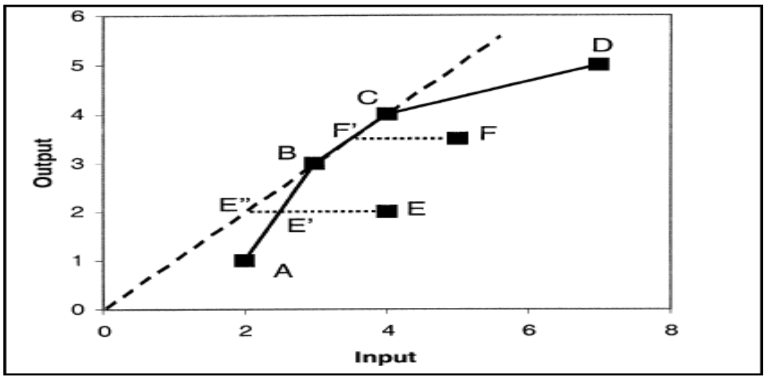


Figure 20. Illustration of DEA Based On a One-Input/Output Case (adopted from Korhonen and Syrjänen, 2003)

Finnish DEA Benchmarking

In practice, the DEA inputs and outputs vary in countries that have implemented this benchmarking method (Jamasb and Pollitt, 2001). In Finland, the EMA has chosen Operating Costs as input, while outputs consist of Interruption Time and Distributed Energy, in addition to two environmental factors (Network Length and Number of Users). The Finnish EMA has used the DEA-model as a part of the monitoring system in regulation of distribution pricing (Korhonen and Syrjänen, 2003). A mathematical presentation of the Finnish DEA model is shown in (7.2), as introduced by Banker and Morey (1986).

$$\text{Min } \theta \tag{7.2}$$

$$\begin{aligned} \text{s.t.} \quad & \theta x_{jo} - \textstyle\sum_{i=1}^{n} \lambda i x_{ji} \geq 0 \;\; (j = 1, \ldots, p), \\ & z_{mo} - \textstyle\sum_{i=1}^{n} \lambda i z_{mi} \geq 0 \;\; (m = 1, \ldots, s), \\ & \textstyle\sum_{i=1}^{n} \lambda i y_{ki} \geq y_{ko} \;\; (k = 1, \ldots, r), \\ & \textstyle\sum_{i=1}^{n} \lambda i = 1 , \\ & \lambda i \geq 0 \qquad\qquad (i = 1, \ldots, n). \end{aligned}$$

where :

θ: efficiency score for the assessed unit,

y_{ki}: observed output k of unit i,

x_{ji}: discretionary input j of unit i (controllable),

z_{mi}: non-discretionary input m of unit i (non-controllable),

λ_i: weight of unit i in the reference point for the assessed unit,

subscript o: refers to the assessed unit, but the unit also preserves its original index.

It is important to note that regardless of which measurements are designated as inputs or outputs, the DEA benchmarking outcome is simply a score between 0 and 1. It cannot provide a comprehensible connection between the efficiency performance and the respective operating circumstances for one to investigate further. Therefore, in this study the SOM is used to visualize the efficiency performance in relation to selected measurements.

7.2 The SOM Model

Data

The data used in this study are derived from the Finnish EMA data sources Tehokkuusluvut 2001-2004 (http://www.energiamarkkinavirasto.fi/). There are in total 356 DSO samples included for training, with 94 in 2001, 90 in 2002, and 86 in 2003 and 2004 respectively. The variables in the SOM-model follow the Finnish DEA efficiency measurements, in order to make the analysis compatible. They are defined as in Korhonen and Syrjänen (2003):

- *Operating Costs (Cost)* – the amount of operating costs (teuro), controlled by the company;
- *Interruption Time (Interruption)* – three year average of customers' total interruption time (h /year);
- *Distributed Energy (Energy)* – the amount of distributed energy weighted by the average national voltage-level-based distribution prices (teuro);
- *Network Length (Length)* – the total network length of the different voltage levels (km);
- *Number of Users (Users)* – the total number of customers.

Model Construction

The map was trained by using four years of data (2001-2004), in order to generate a general picture of this regulatory period. Eight clusters (A-G) were identified on the final SOM display as in Figure 21. This clustering result intuitively reveals the DSOs' geographical operating differences. For instance, the urban based DSOs in clusters B and E have relatively higher Operating Costs, Distributed Energy, and Number of Users, but shorter Network Length and Interruption Time. This is logical as in the urbanized area the population density, the energy demand, and uninterrupted electricity supply requirement are high, whereas the distribution network is concentrated and most likely favors underground cables over overhead lines, which in turn reduces the interruption probability by severe weather. For the DSOs in cluster C which operate in the central lake and north-eastern area, it shows opposing characteristics, i.e., longer Network Length and Interruption Time, but lower user numbers.

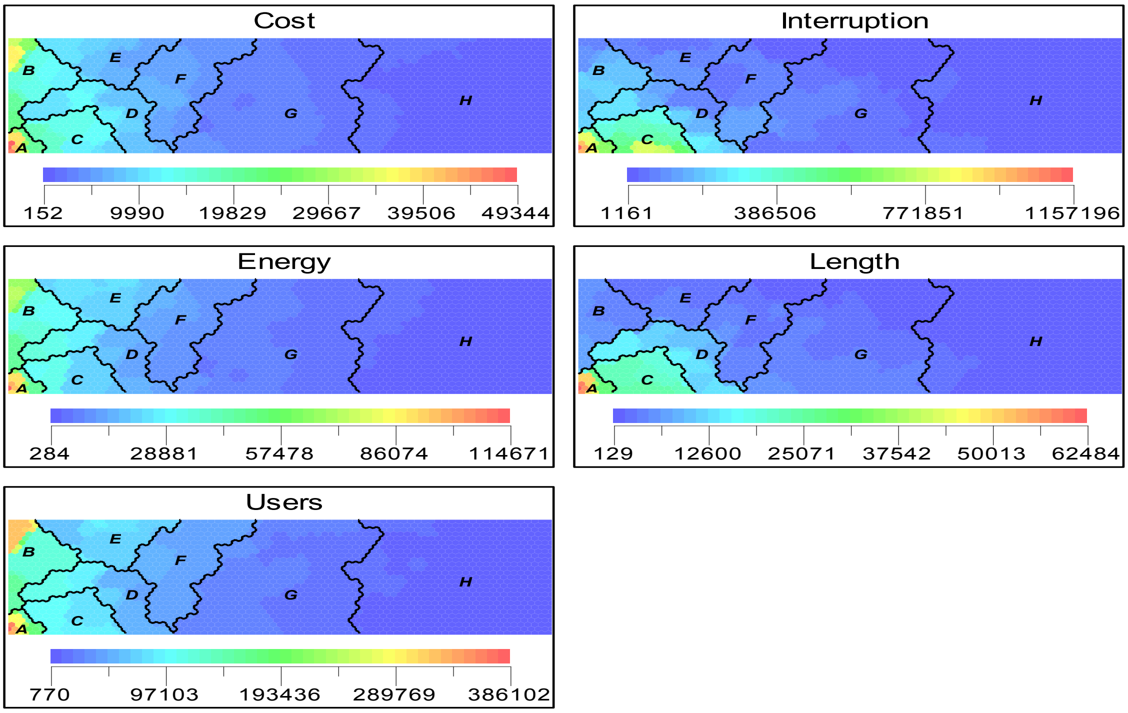


Figure 21. General Picture of Regulatory Period 2001-2004

Then the efficient DSOs[10] (whose efficiency score equals 1) based on the DEA benchmarking were projected onto the general map year by year, in order to visualize the efficiency performance of the DSOs, as shown in Figure 22-25.

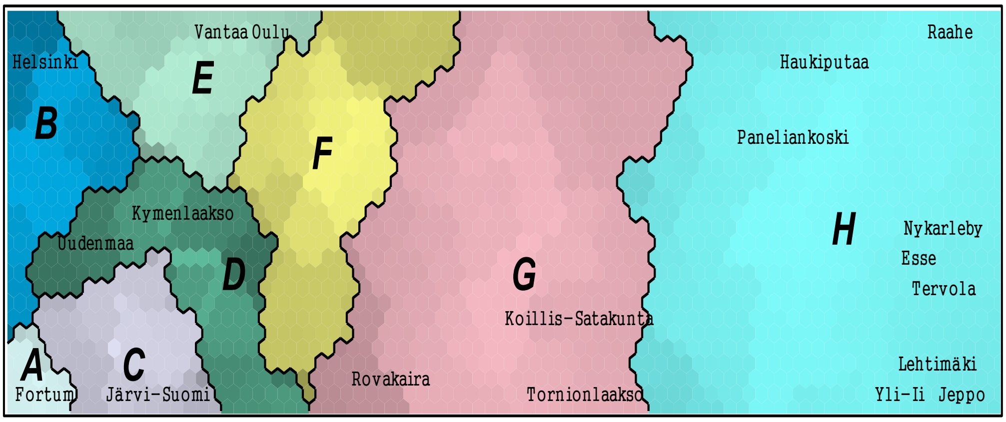


Figure 22. Efficient DSOs in 2001

[10] As many of the Finnish DSOs are municipality based, the company names are simplified accordingly. For example, Helsinki refers to Helsingin Energia, while Tampere stands for Tampereen Sähkölaitos.

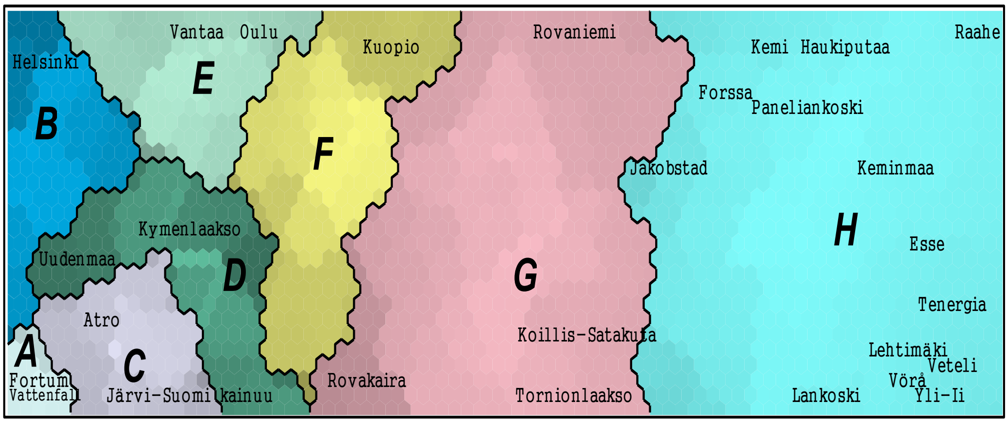


Figure 23. Efficient DSOs in 2002

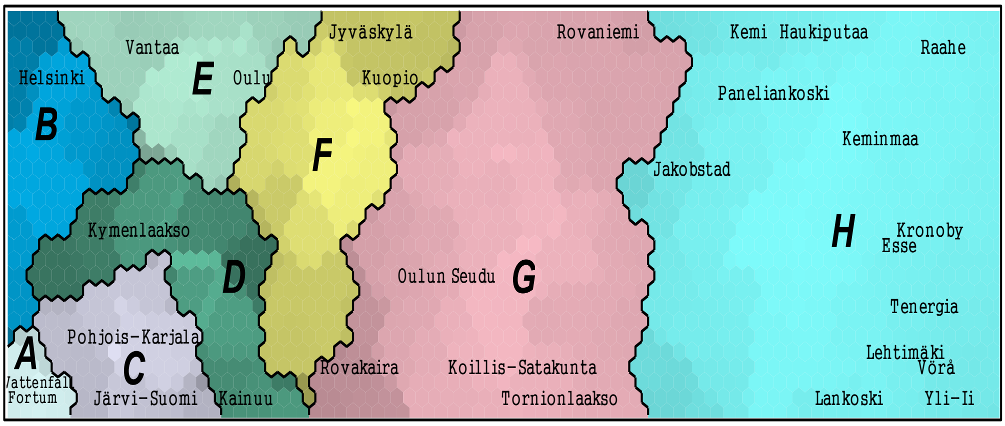


Figure 24. Efficient DSOs in 2003

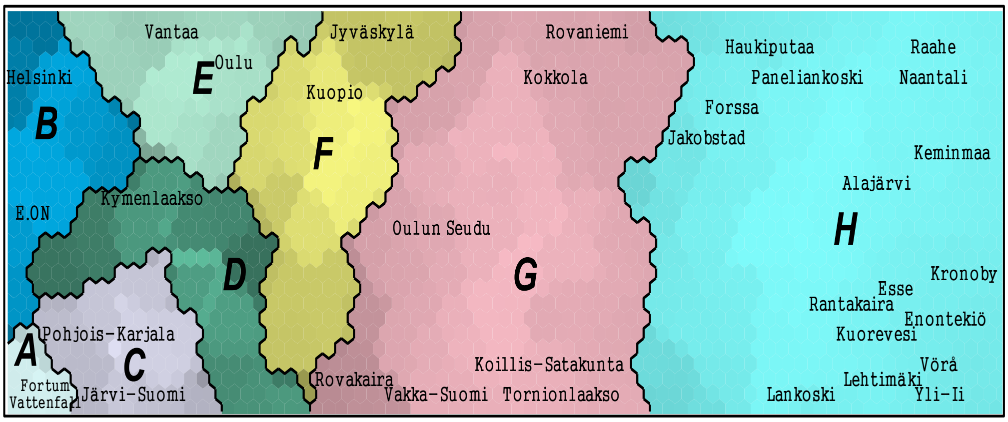


Figure 25. Efficient DSOs in 2004

Visualizing Efficient Performance

Figure 22-25 showed that a great number of DSOs retained efficient performance stably from 2001 to 2004, such as Fortum in cluster A, Helsinki in cluster B, Järvi-Suomi in cluster C, Kymenlaakso in cluster D, Vantaa and Oulu in cluster E, Koillis-Satakunta, Rovakaira, and Tornionlaakso in cluster G, as well as Esse, Haukiputaa, Lehtimäki, Paneliankoski, Raahe, and Yli-Ii in cluster H.

Also, the maps illustrate the changes in each cluster across 2001-2004, with more DSOs becoming efficient. For instance, the number of efficient DSOs in Cluster F, G, and H increased from 0, 3, and 9 DSOs in 2001 (see Figure 22) to 2, 7, and 17 DSOs in 2004 (see Figure 25), respectively. This implies that there has been a considerable efficiency improvement in the Finnish electricity distribution sector. One can as well identify which particular DSOs improved their efficiency, such as E.ON, Jyväskylä, Rovaniemi, Jakobstad, Kemi, etc. The maps also reveal the unstable performance of certain DSOs. For instance, Atro from cluster C (2002) and Kainuu from cluster D (2002 and 2003) appear as efficient, but only for those years.

7.3 Implications

This SOM-based data mining application on distribution efficiency performance visualization provides an alternative approach for understanding the distribution sector. Several implications can be drawn:

Firstly, based on domain experts' evaluation, such a clustering result represents the operating differences of the Finnish DSOs. In other words, using the SOM can visualize the operational environment of DSOs for comparison, instead of only based on the efficiency score. Through cluster analysis, the characteristics of DSOs in a particular group can be discriminated. Not only are the similarities and differences between clusters visible, but the correlations between specific measurements are also displayed. The SOM-model thus offers an opportunity to study the electricity distribution sector on a comparable scale.

Secondly, through visualization, the SOM-model makes the changes and trends in terms of efficiency performance and improvements in the Finnish context explicit. Additionally, it provides the possibility to investigate the efficiency performance in association with the DEA benchmarking measurements in a comprehensible and easy-to-use way. It implies the applicability of the SOM as a complementary analytical tool in electricity distribution benchmarking.

Thirdly, on a related note, the SOM-model would in particular make regulatory sense in connection with the yardstick scheme, used e.g., by Sweden and Spain. In yardstick regulation the goal is to identify a reference performance for utilities to compare with in terms of operating circumstances. The application of the SOM in this study demonstrates its capability to identify the DSOs with similar operating circumstances. Therefore, yardstick regulation may benefit from the application of the SOM for clustering of DSOs. This implies the added value of applying visual data mining with the SOM as a complementary approach in electricity distribution regulation.

It is necessary to note that the constructed SOM-model is based on the Finnish DEA-model measurements. It does not take into account factors such as capital cost and investments. In fact, such factors have important impacts in connection with regulation, benchmarking, and business strategies. It is of interest to include these factors into the SOM approach in future research.

7.4 Chapter Summary

The chapter presented a SOM-based BI application for visualizing efficiency performance in the context of Finnish electricity distribution regulation and efficiency benchmarking. The data investigated are derived from the EMA database for the period of 2001-2004. The variables inspected are in accordance with the Finnish DEA efficiency benchmarking measurements. A SOM-model has been built, which makes the efficiency performance visually comprehensible and comparable. This approach expands the empirical knowledge of BI research in the application domain. As such, this chapter addressed **RO3** and **SO4**.

Chapter 8

Conclusion

"And the end of all our exploring
Will be to arrive where we started
And know the place for the first time."
-- T. S. Eliot, *Little Gidding*

As the title suggests, this thesis aims at advancing BI in facilitating the understanding of customer consumption behavior and residential demand response for the electricity markets of tomorrow. It is an essential task with respect to the overall success of AMI investments and for developing a well-functioning and efficient electricity market. It is a challenging task, given the complex nature of electricity markets, characterized by physical constraints, regulatory policies, the transition to a low carbon society, etc., as well as the challenges of handling massive volumes of real-time smart meter data. This thesis has mainly addressed the *analysis latency* from the perspective of retail pricing, including *innovative retail pricing* and *efficiency benchmarking*. With respect to innovative retail pricing, the research focus is on consumption behavior profiling, real-time BI architecture, price-responsive demand modeling and simulation. With respect to efficiency benchmarking, the research focus is on performance visualization. The thesis has investigated four aspects of BI, namely (i) descriptive analytics based on visual data mining for consumption profiling, (ii) holistic real-time BI enterprise architecture for enhancing management's capacity on real-time decision-making, (iii) prescriptive analytics based on ABM for price-responsive demand simulation, and (iv) a visualization application for electricity distribution benchmarking.

This research is conducted in accordance with the Design Science Research (DSR) paradigm in the IS field. The entire research is constructed through the following steps, with relevance-, rigor-, and design- cycles iterations.

(1) identifying the business opportunities and challenges that are enabled by the implementation of AMI in the problem domain;
(2) defining real-time BI, based on descriptive analytics, prescriptive analytics, and visualization, as solution objective;
(3) designing and developing the data-mining model, the real-time BI framework, and the agent-based modeling and simulation;

(4) demonstrating the real-time BI solution;
(5) performing technical evaluation (model verification) during respective processes of data mining and agent-based modeling;
(6) communicating through peer-reviewed publications and presenting to industry professionals.

This chapter summarizes what has been achieved in the current doctoral studies, the limitations, and what could be carried further in future research.

8.1 What Has Been Achieved

To answer this question, the focus is on the following two aspects:

- Have the research objectives been reached?
- What is the scientific contribution?

Have the research objectives been reached?

In reviewing the research objectives set in the beginning and the studies carried out, a relationship diagram could be drawn, as in Figure 26.

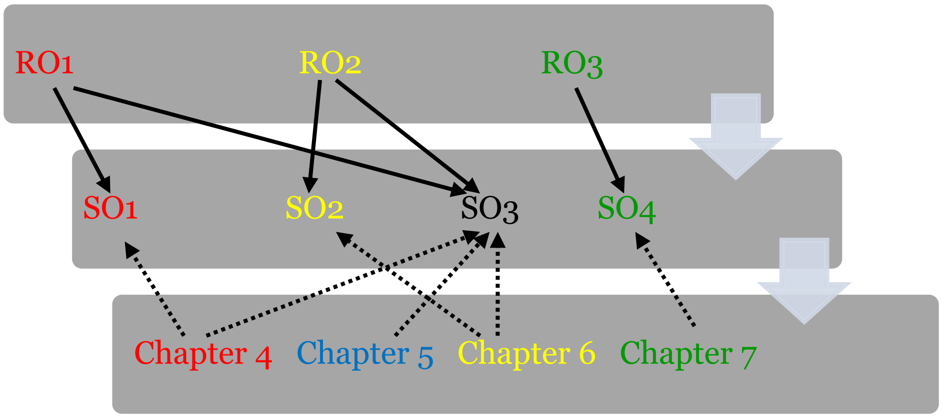


Figure 26. Relation between Research Objectives and Chapters

RO1: to use visual data mining techniques to perform customer consumption behavior profiling;

RO2: to develop a real-time BI framework and a price-responsive demand modeling method for dynamic pricing and demand response applications;

RO3: to visualize the efficient performance of DSOs.

Box 1. Research Objectives

SO1. To explore the potential value of the Self-Organizing Map (SOM) method in electricity consumption profiling;

SO2. To investigate the potential of Agent-based Modeling (ABM) in price-responsive demand simulation;

SO3. To demonstrate how a real-time BI approach can contribute to obtaining and maintaining an innovative and demand response-oriented electricity retail market in the long run;

SO4. To examine the utility of using the Self-Organizing Map (SOM) for efficiency benchmarking.

Box 2. Research Sub-objectives

Chapter 4 investigated the potential value of the SOM method for electricity consumption profiling. The research results indicate the feasibility of the SOM with respect to visual clustering in the problem domain. The studies also explain the added value of applying the SOM as a descriptive analytics approach in advancing a better understanding of customers' electricity consumption patterns and in supporting decision-making on retail pricing. Chapter 4 addresses research objectives **RO1** and **SO1**.

Chapter 5 presented an enterprise-wise real-time BI architecture for enhancing business innovations in the electricity supply business, especially with respect to turning smart meter data into pricing intelligence in the demand response electricity retail market. We argue that real-time BI enterprise architecture is critical in terms of reducing *data latency*, *analysis latency*, and *decision latency*. If empowered by the real-time BI enterprise architecture, management's capacity to maintain sustainable and profitable growth could potentially be increased, given the undergoing complex changes and challenging future of electricity markets. Chapter 5 addresses research objective **RO2**.

Chapter 6 explored the agent-based modeling method for price-responsive demand simulation. The results suggest that as a prescriptive analytics approach, ABM creates a platform for electricity suppliers (retailers) to simulate various scenarios of margin vs. demand changes in relation to spot price. It demonstrates that the ABM approach has potential in developing innovative dynamic pricing and demand response applications. Chapter 6 addresses research objectives **RO2** and **SO2**.

Chapters 4, 5 & 6 together demonstrate that the three BI elements –visual data mining based descriptive analytics, ABM based prescriptive analytics, and holistic real-time BI enterprise architecture – when combined, will contribute to obtaining and maintaining an innovative and demand responsive electricity retail market. Chapters 4, 5 & 6 contribute to **SO3**.

Chapter 7 examined the utility of using the SOM to visualize the efficiency of DSOs for electricity distribution benchmarking in the Finnish context. The SOM-based visualization application makes the efficiency performance visually comprehensible and comparable in terms of operating circumstances. This implies that applying visual data mining with the SOM as a complementary analytical tool is beneficial in electricity distribution regulation and benchmarking. Chapter 7 addresses research objectives **RO3** and **SO4**.

Contribution Claims

Research contributions can be identified as follows:

With respect to real-world practice, firstly, the consumption profiling studies provide actionable insights for the case company in terms of pricing differentiation and targeted retailing. Secondly, the proposed real-time BI framework can be considered by the electricity supply business in the effort to develop demand response electricity retail market. Thirdly, the distribution efficiency visualization study hinted that the yardstick regulatory scheme might benefit from using the SOM-based visual clustering approach.

With respect to the research community, firstly, the current research supplements BI research through examining the four perspectives, namely descriptive analytics based on visual data mining, real-time BI architecture, prescriptive analytics based on agent-based modeling and simulation, and a SOM-based BI application. Secondly, by following the DSR research paradigm, the BI constructs are built in light of rigor and relevance considerations, which will add empirical knowledge to the DSR research in particular and the IS research in general.

With respect to knowledge transferability, exploring timely measured smart meter data for advanced BI applications could induce further scientific collaboration regarding this emerging problem domain, for example, in terms of the intersection between ubiquitous computing, data mining, and demand response simulation for developing interactive mobile applications. It could also contribute to future practice in the energy industry in terms of integrating data mining and agent-based modeling into their pricing decision-making support.

8.2 Limitations and Future Research

This thesis investigated four aspects of BI in relation to the electricity retail markets. The limitations and future research will be discussed in association with each of the aspects:

- Regarding consumption profiling, the smart meter data used in the studies is unfortunately not hourly measured, due to the technical constraints of the meter and the communication technology adopted. Hence, consumption profiling in terms of hourly smart meter data cannot be conducted in this thesis due to data availability. However, using the SOM for streaming-in smart meter data clustering, i.e., the feasibility and performance of the SOM at run-time, is of interest in future research.

- Regarding the real-time BI framework, the implementation and the validation of the real-time BI architecture require field testing. However, the field testing is beyond the scope of the current doctoral study. It would be desirable to have the real-time BI architecture implemented in order to study the impacts in the future.

- Regarding the agent-based price-responsive demand modeling and simulation, the study demonstrated a simplified ad hoc case, i.e., how daily load patterns can be flattened with consumers in a high consumption cluster adjusting their electricity usage according to the spot market price. The simulation experiment mainly observed the interactions between the electricity supplier and the consumers. It did not take into account the actions of other market participants such as the producers and the impacts of external environmental factors except the outdoor temperature. We have chosen the UPPAAL timed automata to formalize the domain model and the UPPAAL tool to run the simulation experiments. As large scale simulation presumes scalable modeling environment, the scalability of this construct needs further examination. Additionally, involving the agent's capacity of learning and adaptation could be the next step in studying the ABM in price-responsive demand simulation.

- Regarding electricity distribution performance visualization, the SOM-based approach is not actually for benchmarking the performance of the DSOs. Rather, it is to provide supplementary information for comparison (ex post) in the Finnish case, though it could be applied for ex ante comparison in other regulatory scheme.

- Last but not least, within the DSR paradigm, evaluation of the design artifacts is required. In the current research, technical evaluation (model verification) is performed. In principle, the technical evaluation and demonstration functioned as proof-of-concept. Further proof-of-value and proof-of-use analyses require field testing, which could be the candidate of future research.

In my view, the current research can be considered as a stepping stone in the effort to advancing BI in the electricity retail markets in terms of consumption profiling and demand response simulation, and a proof-of-concept of the BI artifacts investigated. It is certain that new challenges and research interests will appear along the way.

Bibliography

Abadi, D., Boncz, P., Harizopoulos, S. 2009. Column-oriented Database Systems. In: *Proceedings of the VLDB Endowment* 2(2), pp. 1664– 1665.

Abu-Mostafa, Y. S., Magdon-Ismail, M., Lin, H-T. 2012. *Learning from Data*. AMLBook, US.

Albadi, M. H., El-Saadany, E. F., 2007. Demand Response in Electricity Markets: An Overview. In: *Proceedings of IEEE PES General Meeting (GM' 07)*, Tampa, FL, USA, June 24-28, 2007, pp. 1-5.

Anand, S., Buchner, A. 1998. *Decision Support Using Data Mining*. Financial Times Pitman Publishers, London.

Anderson-Lehman, R., Watson, H. J., Wixom, B. H., Hoffer, J. A. 2004. Continental Airlines Flies High With Real-Time Business Intelligence, *MIS Quarterly Executive* 3(4), pp. 163-176.

Arrow, H., McGrath, J. E., Berdahl, J. L. 2000. *Small Groups as Complex Systems: Formation, Coordination, Development and Adaptation*. Thousand Oaks, CA: Sage Publications.

Axelrod, R. 1997. *The Complexity of Cooperation*. Princeton, NJ: Princeton University Press.

Axelrod, R., Tesfatsion, L. 2006. A Guide for Newcomers to Agent-Based Modeling in the Social Sciences. Appendix, in Tesfatsion L. and Judd K. L. (Eds.), *Handbook of Computational Economics: Volume 2, Agent-Based Computational Economics*, Handbooks in Economics Series, North-Holland, Amsterdam, the Netherlands.

Baars, H. and Zimmer, M. 2013. A Classification for Business Intelligence Agility Indicators. In: *Proceedings of 21st European Conference on Information Systems (ECIS 2013)*, Utrecht, the Netherland, June 5-8, Paper 163.

Bagnall, A., Smith, G., 2005. A multi-agent model of the UK market in electricity generation. *IEEE Transaction on Evolutionary Computation* 9(5), pp. 522–536.

Bailey, J. 2000. Load Profiling for Retail Choice: Examining a Complex and Crucial Component of Settlement. *The Electricity Journal*, December 2000, Elsevier.

Banker, R. D., Morey, R. C. 1986. Efficiency analysis for exogenously fixed inputs and outputs. *Operations Research* 34, pp. 513–521.

Back, B., Sere, K., Vanharanta, H. 1998. Managing Complexity in Large Data Bases Using Self-Organizing Maps. *Information and Organization* 8(4), pp. 191–210, Elsevier.

Belonogova, N., Kaipia, T., Lassila, J., Partanen, J., 2011. Demand Response: Conflict between Distribution System Operators and Retailer. In: *Proceedings of 21st International Conference on Electricity Distribution (CIRED 2011)*, Paper 1085, June 6-9, Frankfurt, Germany.

Benbasat, I., Zmud, R. W. 1999. Empirical Research in Information Systems: The Practice of Relevance. *MIS Quarterly* 23(1), pp. 3-16.

Benbasat, I., Zmud, R. W. 2003. The Identity Crisis within the IS Discipline: Defining and Communicating the Discipline's Core Properties. *MIS Quarterly* 27(2), pp. 183-194.

Bengtsson, J., Yi, W., 2004. Timed Automata: Semantics, Algorithms and Tools. *Lecture Notes on Concurrency and Petri Nets*. W. Reisig and G. Rozenberg (eds.), LNCS 3098, Springer-Verlag, 2004, pp. 87-124.

Bishop, C. M. 1995. *Neural Networks for Pattern Recognition*. Oxford University Press, Avon.

Borrill, P. L., Tesfatsion, L. 2011. Agent-Based Modeling: The Right Mathematics for the Social Sciences? In: Davis, J. B. and Hands D.

W., (Eds.) *The Elgar Companion to Recent Economic Methodology*. Edward Elgar Publishers.

Bower, J., Bunn, D.W., 2000. Model-based comparisons of pool and bilateral markets for electricity. *Energy Journal* 21(3), pp. 1–29.

Bower, J., Bunn, D.W., Wattendrup, C., 2001. A model-based analysis of strategic consolidation in the German electricity industry. *Energy Policy* 29(12), pp. 987–1005.

Cabena, P., Hadjinian, P., Stadler, R., Verhees, J., Zanasi, A. 1998. *Discovering Data Mining: From Concepts to Implementation*. Prentice Hall Saddle River, New Jersey.

Cai, C-R., Jahangiri, P., Thomas, A. G., Zhao, H., Aliprantis, D. C., Tesfatsion, L. 2011. Agent-Based Simulation of Distribution Systems with High Penetration of Photovoltaic Generation. In: *Proceedings of the IEEE Power and Energy Society General Meeting*, Detroit, MI, July 24-28.

Cau, T. D. H., Anderson, E. J. 2002. A co-evolutionary approach to modelling the behaviour of participants in competitive electricity markets. In: *Proceedings of IEEE Power Engineering Society Summer Meeting* 3, pp. 1534–1540.

CEER (Council of European Energy Regulators) 2013. Status Review of Regulatory Aspects of Smart Metering, Ref: C13-RMF-54-05, 12 September 2013.

Charnes, A., Cooper, W. W., Rhodes, E. 1978. Measuring the efficiency of decision making units. *European Journal of Operational Research* 2, pp. 429–444.

Chaudhuri, S., Dayal, U., Narasayya, V. 2011 An overview of business intelligence technology. *Communication of the ACM* 54(8), pp. 88-98.

Chicco, G., Napoli, R., Postolache, P., Scutariu, M., Toader, C. 2003. Customer characterization options for improving the tariff offer. *IEEE Transactions on Power Systems* 18(1), pp. 381–387.

Cios, K., Teresinska, A., Konieczna, S., Potocka, J., Sharma, S. 2000. Diagnosing myocardial perfusion from SPECT bull's-eye maps – a knowledge discovery approach. *IEEE Engineering in Medicine and Biology Magazine, special issue on Medical Data Mining and Knowledge Discovery* 19(4), pp.17–25.

Cios, K., Kurgan, L. 2005. Trends in data mining and knowledge discovery. In: Pal, N.R., and Jain L.C. (Eds.), *Advanced Techniques in Knowledge Discovery and Data Mining*, 1–26, Springer-Verlag, London.

Collica, R. S. 2007. *CRM Segmentation and Clustering Using SAS Enterprise Miner*. SAS Publishing.

Comnes, G. A., Stoft, S., Greene, N., Hill, L. J. 1995. Performance-Based Ratemaking for Electric Utilities. *Review of Plans and Analysis of Economic and Resource Planning Issues*, Vol. I. Oak Ridge, Tenn. Oak Ridge National Laboratory and University of California, Berkeley.

Davenport, T. H., Harris, J. G. 2007. *Competing on Analytics: The New Science of Winning*. Harvard Business School Press, Boston.

Deboeck, G., Kohonen, T. 1998. *Visual explorations in finance using self-organizing maps*. Springer, Berlin.

Denny, Willianms, G. J., Christen, P. 2010. Visualizing Temporal Cluster Changes Using Relative Density Self-Organizing Maps. *Knowledge and Information Systems* 25(2), pp. 281-302.

Drazin, R., Sandelands, L. 1992. Autogenesis: A Perspective on the Process of Organizing. *Organization Science* 3(2), pp. 230-249.

Edvardsen, D. F., Førsund, F. R. 2003. International benchmarking of electricity distribution utilities. *Resource and Energy Economics* 25, pp. 353–371.

Ehlen, M. A., Scholand, A. J., Stamber, K. L., 2007. The effects of residential real-time pricing contracts on transco loads, pricing, and profitability: Simulations using the N-ABLE™ agent-based model. *Energy Economics* 29(2), Elsevier, March 2007, pp. 211–227.

Ekici, S., 2009. Classification of power system disturbances using support vector machines. *Expert Systems with Applications* 36(6), August 2009, pp. 9859–9868, Elsevier.

Eklund, T., Back, B., Vanharanta, H., Visa, A. 2008. Evaluating a SOM-Based Financial Benchmarking Tool. *Journal of Emerging Technologies in Accounting* 5(1), pp. 109–127.

Epstein, J. M., Axtell, R. 1996. *Growing Artificial Societies: Social Science from the Bottom Up*. Washington, DC: Brookings Institution Press.

Ericson, T. 2007. *Short-term electricity demand response*. PhD thesis. ISBN 978-82-471-1156-7.

Ernst, D., Glavic, M., Wehenkel, L. 2004. Power systems stability control: reinforcement learning framework. *IEEE Transactions on Power Systems* 19(1), pp. 427-435.

Espinoza, M., Joye, C., Belmans, R., DeMoor, B. 2005. Short-term load forecasting, profile identification, and customer segmentation: a methodology based on periodic time series. *IEEE Transactions on Power Systems* 20(3), pp.1622–1630.

Etzion, O., Niblett, P. 2010. *Event Processing in Action*. Manning Publications, Shelter Island, New York.

European Technology Platform SmartGrids 2010. Strategic Deployment Document for Europe's Electricity Networks of the Future, April, 2010.

Evans, J. R. 2013. *Business analytics: Methods, models, and decisions*. Prentice Hall.

Farsi, M., Fetz, A., Filipini, M. 2007. Benchmarking and regulation in the electricity distribution sector. *CEPE Working Paper No. 54, Centre for Energy Policy and Economics*. Swiss Federal Institutes of Technology, Zürich.

Faruqui, A., Wood, L. 2008. *Quantifying the Benefits of Dynamic Pricing in the Mass Market*. Edison Electric Institute. January, 2008.

Fayyad, U., Piatesky-Shapiro, G., Smyth, P., Uthurusamy, R. (Eds.), 1996. *Advances in Knowledge Discovery and Data Mining*. AAAI Press, Cambridge.

Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. 1996. From data mining to knowledge discovery: an overview. In: Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., and Uthurusamy, R. (Eds.), *Advances in Knowledge Discovery and Data Mining*, pp. 1–34, AAAI Press, Cambridge.

Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. 1996. The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM* 39(11), pp. 27–34.

Ferreira de Oliveira, M. C., Levkowistz, H. 2003. From visual data exploration to visual data mining: a survey. *IEEE Transactions on Visualization and Computer Graphics* 9(3), pp. 378-394.

Ferreira, J., Ramos, S., Vale, Z., Soares, J. 2011. A Data-Mining-Based Methodology for Transmission Expansion Planning. *IEEE Intelligent Systems*, March/April, pp. 28-37.

Figueiredo, V., Rodrigues, F., Vale, Z., Gouveia, J. 2005. An electric energy consumer characterization framework based on data mining techniques. *IEEE Transactions on Power Systems* 20(2), pp. 596–602.

Gell-Mann, M. 1994. *The Quark and the Jaguar: Adventures in the Simple and the Complex*. New York: W. H. Freeman.

Gregor, S. 2006. The Nature of Theory in Information Systems. *MIS Quarterly* 30(3), pp. 611-642.

Gregor, S., Jones, D. 2007. The Anatomy of a Design Theory, *Journal of the Association for Information Systems* 8(5), pp. 313-335.

Gregor, S., Hevner, A. R. 2013. Positioning Design Science Research for Maximum Impact. *MIS Quarterly* 37(2), pp. 337-355.

Guerci, E., Rastegar, M. A., Cincotti, S. 2010. Agent-Based Modeling and Simulation of Competitive Wholesale Electricity Markets. *Handbook of Power Systems II*, Springer, pp. 241-286.

Hackathorn, R. 2004. The BI Watch: Real-Time to Real Value, *DM Review* 14(1), pp. 24-43.

Hall, G. L. 2000. Electricity Pricing and Regulatory Practice in a Competitive Environment. *Workshop Paper No. 2: Analysis of Alternative Ratemaking Methodologies*. Manila. Prepared by Hagler Bailly Services, Inc. for the Asian Development Bank and Philippines Energy Regulatory Board.

Han, J., Kamber, M. 2000. *Data Mining: Concepts and Techniques*. The Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann.

Haykin, S. 1999. *Neural networks: a comprehensive foundation, 2nd ed*. Prentice Hall, Upper Saddle River.

Hevner, A. R., March, S. T., Park, J., Ram, S. 2004. Design Science in Information Systems Research. *MIS Quarterly* 28(1), pp. 75-105.

Hevner, A. 2007. A Three Cycle View of Design Science Research. *Scandinavian Journal of Information Systems* 19(2), pp. 87-92.

Hill, L. 1995. *A Primer on Incentive Regulation for Electric Utilities*. Oak Ridge National Laboratory, ORNL/CON-422.

Holland, J. H. 1995. *Hidden Order: How Adaptation Builds Complexity*. Reading, MA: Perseus Books.

Holland, J. H. 2006. Studying Complex Adaptive Systems. *Journal of Systems Science and Complexity* 19(1), pp. 1-8.

Holmbom, A. H., Eklund, T., Back, B. 2011. Customer Portfolio Analysis Using the SOM. *International Journal of Business Information Systems* 8(4), pp. 396–412.

Honkapuro, S., Lassila, J., Viljainen, S., Tahvanainen, K., Partanen, J. 2004. Effects of benchmarking of electricity distribution companies in Nordic countries – comparison between different benchmarking methods. In: *Proceedings of Nordic Distribution and Asset Management Conference (Nordac 2004)*, Espoo, Finland, August 23-24.

Hämäläinen, R. P., Mäntysaari, J., Ruusunen, J., Pineau, P. O., 2000. Cooperative consumers in a deregulated electricity market — dynamic consumption strategies and price coordination. *Energy* 25(9), Elsevier, September 2000, pp. 857–875.

Iivari, J. 2003. The IS Core – VII: Towards Information Systems as a Science of Meta-Artifacts. *Communications of the Association for Information Systems* 12, pp. 568-581.

Iivari, J. 2007. A Paradigmatic Analysis of Information Systems as a Design Science. *Scandinavian Journal of Information Systems* 19(2), pp. 39-63.

Jamasb, T., Pollitt, M. 2000. Benchmarking and regulation of electricity transmission and distribution utilities: lessons from international experience. In: *Cambridge Working Papers in Economics 0101*, Faculty of Economics, University of Cambridge.

Jamasb, T., Pollitt, M. 2001. Benchmarking and regulation: international electricity experience. *Utilities Policy* 9(3), pp. 107–130.

Jamasb, T., Pollitt, M. 2003. International benchmarking and regulation: an application to European electricity distribution utilities. *Energy Policy* 31(15), pp. 1609–1622.

Johnsen, T. A., Olsen, O. J. 2011. Regulated and Unregulated Nordic Retail Prices. *Energy Policy* 39(6), pp. 3337-3345, Elsevier.

Joskow, P. J., Schmalensee, R. 1986. Incentive regulation for electric utilities. *Yale Journal on Regulation* 4(1), pp. 1–49.

Jourdan, Z., Rainer, R. K., Marshal, T. E. 2008. Business Intelligence: an Analysis of the Literature. *Information Systems Management* 25, pp. 121-131.

Keim, D. 2001. Visual exploration of large data sets. *Communications of the ACM* 44(8), pp. 38-44.

Keppo, J., Räsänen, M. 1999. Pricing of electricity tariffs in competitive markets. *Energy Economics* 21, pp. 213–223.

Keshav, S., Rosenberg, C. 2011. How internet concepts and technologies can help green and smarten the electrical grid. *ACM SIGCOMM Computer Communication Review* 41(1), pp. 109–114.

Kim, H. S., Mun, K. J., Lee, H. S. 2000. Electric load forecasting and classification using neural networks for the EMS. In: *Proceedings of the IASTED International Conference Power and Energy Systems*, IASTED/ACTA Press, Anaheim, CA, USA, pp. 171–175.

Kinnunen, K. 2005. Pricing of electricity distribution: an empirical efficiency study in Finland, Norway and Sweden. *Utilities Policy* 13, pp. 15–25.

Kinnunen, K. 2006. Investment incentives: regulation of the Finnish electricity distribution. *Energy Policy* 34, pp. 853–862.

Kiviluoto, K. 1996. Topology Preservation in Self-Organizing Maps. In: *Proceedings of the IEEE International Conference on Neural Networks*, Washington, DC, June 3-6, pp. 294-299.

Koesrindartoto, D., Sun, J., Tesfatsion, L., 2005. An Agent-Based Computational Laboratory for Testing the Economic Reliability of Wholesale Power Market Desings. In: *Proceedings of the IEEE Power Engineering Society General Meeting*, vol. 3, San Francisco, CA, USA, June 12-16, pp. 2818–2823.

Kohonen, T. 2001. *Self-organizing maps, 3rd ed.* Springer, Berlin.

Kok, K., Scheepers, M., Kamphuis, R. 2009. Intelligence in Electricity Networks for Embedding Renewables and Distributed Generation. In: R.R. Negenborn, Z. Lukszo, and J. Hellendoom (Ed.), *Intelligent Infrastructures*, Springer.

Korhonen, P. J., Syrjänen, M. J. 2003. Evaluation of cost efficiency in Finnish electricity distribution. *Annals of Operations Research* 121, pp. 105–122.

Lendasse, A., Lee, J., Wertz, V., Verleysen, M. 2002. Forecasting electricity consumption using nonliner projection and self-organizing maps. *Neurocomputing* 48, pp. 299–311.

Li, H., Tesfatsion, L. 2009. Development of Open Source Software for Power Market Research: The AMES Test Bed. *Journal of Energy Markets* 2(2), pp. 111-128.

Liberatore, M. J., Luo, W. H. 2010. The analytics movement: Implications for operations research. *Interfaces* 40(4), July-August, pp. 313-324.

Luckham, D. 2002. *The Power of Events: An Introduction to Complex Event Processing in Distributed Enterprise Systems*. Addison-Wesley, Boston, MA, USA.

Luckham, D. 2012. *Event Processing for Business: Organizing the Real-Time Enterprise*. John Wiley & Sons, Hoboken, New Jersey.

Maghrabi, R. O., Oakley, R. L., Thambusamy, R., Iyer, L. 2011. The Role of Business Intelligence (BI) in Service Innovation: an Ambidexterity Perspective. In: Proceedings of American

Conference on Information Systems (AMCIS 2011), August 4-7, Detroit, MI, US, Paper 319.

March, S. T., Hevner, A. R. 2007. Integrated Decision Support Systems: A Data Warehousing Perspective. *Decision Support Systems* 43(3), pp. 1031-1043.

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

March, S. T., Storey, V. C. 2008. Design Science in the Information Systems Discipline: An Introduction to the Special Issue on Design Science Research. *MIS Quarterly* 32(4), pp. 725-730.

Marghescu, D. 2006. Evaluating the Effectiveness of Projection Techniques in Visual Data Mining. In: Villanueva J. J. (Eds.), *Proceedings of the 6th IASTED International Conference on Visualization, Imaging, and Image Processing (VIIP2006)*, pp. 94-103, August 28-30, Palma de Mallorca, Spain. ACTA Press.

Marjanovic, O., Roose, R. 2011. BI-Enabled, Human-Centric Business Process Improvement in a Large Retail Company. In: *Proceedings of the 44th Hawaii International Conference on System Sciences(HICSS 2011)*, 4-7 January, 2011, Koloa, Kauai, HI, USA, pp. 1-9.

Miller, J. H., Page, S. E. 2007. *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*. Princeton, NJ: Princeton University Press.

Müller, M., Sensfuß, F., Wietschel, M., 2007. Simulation of current pricing-tendencies in the German electricity market for private consumption. *Energy Policy* 35(8), Elsevier, August 2007, pp. 4283–4294.

Nababhushana, T. N., Veeramanju, K. T., Shivanna. 1998. Coherency identification using growing self-organizing feature maps (power system stability). In: *IEEE Proceedings of*

International Conference on Energy Management and Power Delivery (EMPD 1998), vol. 1, pp. 113–116.

Nan, N. 2011. Capturing Bottom–Up IT Use Processes. *MIS Quarterly* 35(2), pp. 505-532.

Nicolaisen, J., Smith, M., Petrov, V., Tesfatsion, L., 2000. Concentration and capacity effects on electricity market power. In: *Proceedings of 2000 Congress on Evolutionary Computation*, vol. 2, La Jolla, USA, pp. 1041–1047.

Nicolaisen, J., Petrov, V., Tesfatsion, L., 2001. Market power and efficiency in a computational electricity market with discriminatory double-auction pricing. *IEEE Transactions on Evolutionary Computation* 5(5), pp. 504–523.

Nizar, A. H., Dong, Z. Y., Zhao, J. H. 2006. Load profiling and data mining techniques in electricity deregulated market. In: *Proceedings of IEEE Power Engineering Society General Meeting*, Montreal, Canada, June 18-22, pp. 1-7.

Nunamaker, J. F., Chen, M., Purdin, T. D. M. 1990. Systems Development in Information Systems Research, *Journal of Management Information Systems* 7(30), pp. 89-106.

Oja, M., Kaski, S., Kohonen, T. 2002. Bibliography of Self-Organizing Map (SOM) Papers: 1998-2001 Addendum. *Neural Computing Surveys* 3, pp. 1–156.

Olbrich, S., Poeppelbuss, J., Niehaves, B. 2011. BI Systems Managers' Perception of Critical Contextual Success Factors: A Delphi Study. In: *Proceedings of International Conference on Information Systems (ICIS 2011)*, Paper 12, December 4-7, Shanghai, China.

Olivia, R. 2009. *Business Intelligence Success Factors: Tools for Aligning Your Business in the Global Economy*. Hoboken, N.J. Wiley & Sons.

Orlikowski, W. J., Iacono, C. S. 2001. Research Commentary: Desperately Seeking the 'IT' in IT Research—A Call to Theorizing the IT Artifact. *Information Systems Research* 12(2), pp. 121-134.

Peffers, K., Tuunanen, T., Rothenberger, M. A., Chatterjee, S. 2008. A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems* 24(3), pp. 45-77.

Petrov, V., Sheblé, G., 2001. Building electric power auctions with improved Roth-Erev reinforced learning. In: *Proceedings of the 33rd North American Power Symposium*, Texas, USA, October 15-16, pp. 10-13.

Pitt, B. D., Kitschen, D. S. 1999. Application of data mining techniques to load profiling. In: *Proceedings of 21st IEEE international conference on power industry computer applications (PICA '99)*, Santa Clara, CA, May 16-21, pp. 131–136.

Pölzlbauer, G. 2004. Survey and Comparison of Quality Measures for Self-Organizing Maps. In: *Proceedings of the 5th Workshop on Data Analysis*, Tatranska Polianka, Slovak Republic, June 24-27, pp. 67-82.

Ramos, C., Liu, C-C. 2011. AI in power systems and energy markets. *IEEE Intelligent Systems* 26(2), pp. 5–8.

Ramos, S., Vale, Z. 2008. Data mining techniques to support the classification of MV electricity customers. In: *Proceedings of IEEE Power and Energy Society General Meeting*, Pittsburgh, PA, USA, July 20-24, pp. 1–7.

Rehtanz, C. 1999. Visualisation of voltage stability in large electric power systems. In: *IEEE Proceedings Generation, Transmission and Distribution* 146(6), November, pp. 573–576.

Richter, C.W., Sheblé, G.B., 1998. Genetic algorithm evolution of utility bidding strategies for the competitive marketplace. *IEEE Transactions on Power Systems* 13(1), pp. 256–261.

Riqueline, J., Martinez, J. L., Gomez, A., Goma, D. C. 2000. Possibilities of artificial neural networks in short-term load forecasting. In: *Proceedings of the IASTED International Conference Power and Energy Systems*, IASTED/ACTA Press, Anaheim, CA, USA, pp. 165–170.

Rodriguez, C. P., Anders, G. J. 2004. Energy price forecasting in the Ontario competitive power system market. *IEEE Transactions on Power Systems* 19(1), pp. 366-374.

Ronald, S. S. 2001. *Accelerating Customer Relationships: Using CRM and Relationship Technologies*. Prentice Hall PRT, Upper Saddle River, New Jersey.

Roop, J. M., Fathelrahman, E., 2003. Modeling electricity contract choice: an agent-based approach. In: *Proceedings of the ACEEE Summer Study Meeting* vol. 6, Rye Brook, New York, July 29-August 1, pp. 193-204.

Rosemann, M., Vessey, I. 2008. Toward Improving the Relevance of Information Systems Research to Practice: The Role of Applicability Checks. *MIS Quarterly* 32(1), pp. 1-22.

Russell, S., Norvig, P. 2003. *Artificial Intelligence: A Modern Approach 2nd ed.* Upper Saddle River, New Jersey: Prentice Hall.

Saldanha, T., Krishnan, M. 2011. BI and CRM for Customer Involvement in Product and Service Development. In: *Proceedings of International Conference on Information Systems (ICIS 2011)*, Paper 14, December 4-7, Shanghai, China.

Samuel, A. 1959. Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development* 3, pp. 210-229.

Sarlin, P. 2011. Clustering the Changing Nature of Currency Crises in Emerging Markets: An Exploration with Self-Organising Maps. *International Journal of Computational Economics and Econometrics* 2(1), pp. 24-46.

Sawyer, R. K. 2005. *Social Emergence: Societies as Complex Systems*. Cambridge, UK: Cambridge University Press.

Schelling, T. 1978. *Micro Motives and Macro Behavior*. Norton, New York, USA.

Schuler, R. E. 2012. Planning, Markets and Investment in the Electric Supply Industry. In: *Proceedings of 45th Hawaii International Conference on System Sciences (HICSS 2012)*, Maui, Hawaii, USA, January 4-7, pp. 1923-1930.

Serrano-Cinca, C. 1996. Self Organizing Neural Networks for Financial Diagnosis. *Decision Support Systems* 17(3), pp. 227-238.

Shearer, C. 2000. The CRISP-DM model: the new blueprint for data mining. *Journal of Data Warehousing* 5(4), pp. 13–19.

Simon, H. A. 1957. A Behavioral Model of Rational Choice. In: *Models of Man, Social and Rational: Mathematical Essays on Rational Human Behavior in a Social Setting*. New York: Wiley.

Simon, H. A. 1996. *The Science of the Artificial*. Cambridge, USA: The MIT Press.

Sun, J., Tesfatsion, L. 2007. Dynamic testing of wholesale power market designs: an open-source agent-based framework. *Computational Economics* 30(3), pp. 291–327.

Tesfatsion, L. 2006. Agent-Based Computational Economics: A Constructive Approach to Economic Theory. Introductory chapter (pp. 831-880) In: Tesfatsion L. and Judd K. L. (Eds.), *Handbook of Computational Economics:Volume 2, Agent-Based Computational Economics*, Handbooks in Economics Series, North-Holland, Amsterdam, the Netherlands.

Tesfatsion, L., Judd, K. L. 2006. *Handbook of Computational Economics: Volume 2, Agent-Based Computational Economics*, Handbooks in Economics Series, North-Holland, Amsterdam, the Netherlands.

Tesfatsion, L. 2009. Auction Basics for Wholesale Power Markets: Objectives and Pricing Rules. In: *Proceedings of the IEEE Power and Energy Society General Meeting*, Calgary, Alberta, CA, July 26-30.

Thomas, A. G., Cai, C-R., Aliprantis, D. C., Tesfatsion, L. 2012. Effects of Price-Responsive Residential Demand on Retail and Wholesale Power Market Operations. In: *Proceedings of the IEEE Power and Energy Society General Meeting*, San Diego, CA, USA, July 22-26.

Ultsch, A., Siemon, H. 1990. Kohonen's Self Organizing Feature Maps for Exploratory Data Analysis. In: *Proceedings of the International Conference on Neural Networks (ICNN90)*, Dordrecht, the Netherlands, pp. 305-308.

Vale, Z., Morais, H., Ramos, S., Soares, J., Faria, P. 2011. Using Data Mining Techniques to Support DR Programs Definition in Smart Grids. In: *Proceedings of IEEE Power and Energy Society General Meeting*, Detroit, MI, USA, July 24-28, pp. 1-8.

van Aken, J. E. 2004. Management Research Based on the Paradigm of the Design Sciences: The Quest for Field-Tested and Grounded Technological Rule. *Journal of Management Studies* 41(2), pp. 219-246.

Venna, J. Kaski, S. 2001. Neighborhood preservation in nonlinear projection methods: An experimental study. In: *Proceedings of the International Conference on Artificial Neural Networks*, Vienna, Austria, August 21-25.

Verdu, S., Garcia, M., Franco, F., Encinas, N., Marin, A., Molina, A., Lazaro, E. 2004. Characterization and Identification of Electrical Customers Through the Use of Self-Organizing Maps and Daily Load Parameters. In: *Proceedings of IEEE Conference on Power Systems Conference and Exposition (PSCE 2004)* vol. 2, pp. 899-906.

Vesanto, J., Alhoniemi, E. 2000. Clustering of the self-organizing map. *IEEE Transactions on Neural Networks* 11(3), pp. 586-600.

Vesanto, J., Alhoniemi, E., Himberg, J., Kiviluoto, K., Parviainen, J. 1999. Self-Organizing Map for Data Mining in Matlab: the SOM Toolbox. *Simulation New Europe* 25(54), March.

Vesanto, J., Himberg, J., Alhoniemi, E., Parhankangas, J. 2000. *SOM Toolbox for Matlab5*. Helsinki University of Technology, Technical Report A57.

Walls, J. G., Widmeyer, G. R., Sawy, O. A. E. 1992. Building an Information System Design Theory for Vigilant EIS. *Information Systems Research* 3(1), pp. 36-59.

Ward, J. H. Jr. 1963. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association* 58, pp. 236–244.

Watson, H. J., Wixom, B. H. 2007. The Current State of Business Intelligence. *Computer* 40(9), pp. 96-99.

Watson, H. J. 2009. Tutorial: Business Intelligence - Past, Present, and Future. *Communications of the Association for Information Systems* 25(39), pp. 487-510.

Weidlich, A., Veit, D. 2008. A critical survey of agent-based wholesale electricity market models. *Energy Economics* 30, Elsevier, pp. 1728-1759.

Wixom, B. H., Watson, H. J. 2010. The BI-Based Organization. *International Journal of Business Intelligence Research* 1(1), pp. 13-28.

Wooldridge, M., Jennings, N. R. 1995. Intelligent agents: theory and practice. *Knowledge Engineering Review* 10(2), pp. 115-152.

Yao, Z., Eklund, T., Back, B. 2010. Using SOM-Ward Clustering and Predictive Analytics for Conducting Customer Segmentation. In: *Proceedings of 2010 IEEE International Conference on Data Mining Workshops*, pp. 639-646.

Zhang, T., Nuggall, J. W. 2011. Evaluating Government's Policies on Promoting Smart Metering Diffusion in Retail Electricity Markets via Agent-Based Simulation. *Journal of Production Innovation Management* 28, pp. 169-186.

Zhou, Z., Zhao, F., Wang, J. 2011. Agent-Based Electricity Market Simulation with Demand Response from Commercial Buildings. *IEEE Transactions on Smart Grid* 2(4), pp. 580-588.

Zimmer, M., Baars, H., Kemper, H-G. 2012. The Impact of Agility Requirements on Business Intelligence Architectures. In: *Proceedings of 45th Hawaii International Conference on System Sciences (HICSS 2012)*, pp. 4189-4198.

Part II

Original Publications

1

Publication I

Liu, H., Eklund, T., and Back, B. (2012). Smart Metering and Customer Consumption Behavior Profiling: Exploring Potential Business Opportunities for DSOs and Electricity Retailers. In: Jussi Kantola, Waldemar Karwowski (Eds.), Knowledge Service Engineering Handbook, p 179-189, Taylor & Francis, 2012.

Publication II

2

Liu, H., Yao, Z., Eklund, T., and Back, B. (2012). Electricity Consumption Time Series Profiling: A Data Mining Application in Energy Industry. In: Petra Perner (Ed.), Advances in Data Mining: Applications and Theoretical Aspects, LNAI 7377, p 52–66, Springer, 2012.

Publication III

Liu, H., Yao, Z., Eklund, T., and Back, B. (2012). From Smart Meter Data to Pricing Intelligence – Visual Data-Mining towards Real-Time BI. In: K. D. Joshi, Youngjin Yoo (Eds.), AMCIS 2012 proceeding, Paper 11, AISeL, 2012.

3

Publication IV

Liu, H. and Vain, J. (2013). An Agent-Based Modeling for Price-Responsive Demand Simulation. In: Proceedings of 15th International Conference on Enterprise Information Systems (ICEIS 2013), Paper 67.

4

Publication V

Liu, H., Eklund, T., Back, B., and Vanharanta, H. (2011). Visual Data Mining: Using Self-Organizing Maps for Electricity Distribution Regulation. In: Proceedings of International Conference on Digital Enterprise and Information Systems (DEIS 2011), July 20-22, London, UK. Springer, CCIS194, p 631-646.

5

Turku Centre for Computer Science
TUCS Dissertations

1. **Marjo Lipponen**, On Primitive Solutions of the Post Correspondence Problem
2. **Timo Käkölä**, Dual Information Systems in Hyperknowledge Organizations
3. **Ville Leppänen**, Studies on the Realization of PRAM
4. **Cunsheng Ding**, Cryptographic Counter Generators
5. **Sami Viitanen**, Some New Global Optimization Algorithms
6. **Tapio Salakoski**, Representative Classification of Protein Structures
7. **Thomas Långbacka**, An Interactive Environment Supporting the Development of Formally Correct Programs
8. **Thomas Finne**, A Decision Support System for Improving Information Security
9. **Valeria Mihalache**, Cooperation, Communication, Control. Investigations on Grammar Systems.
10. **Marina Waldén**, Formal Reasoning About Distributed Algorithms
11. **Tero Laihonen**, Estimates on the Covering Radius When the Dual Distance is Known
12. **Lucian Ilie**, Decision Problems on Orders of Words
13. **Jukkapekka Hekanaho**, An Evolutionary Approach to Concept Learning
14. **Jouni Järvinen**, Knowledge Representation and Rough Sets
15. **Tomi Pasanen**, In-Place Algorithms for Sorting Problems
16. **Mika Johnsson**, Operational and Tactical Level Optimization in Printed Circuit Board Assembly
17. **Mats Aspnäs**, Multiprocessor Architecture and Programming: The Hathi-2 System
18. **Anna Mikhajlova**, Ensuring Correctness of Object and Component Systems
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 Market 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. **Frank 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
57. **Mikael Collan**, Giga-Investments: Modelling the Valuation of Very Large Industrial Real Investments
58. **Dag Björklund**, A Kernel Language for Unified Code Synthesis
59. **Shengnan Han**, Understanding User Adoption of Mobile Technology: Focusing on Physicians in Finland
60. **Irina Georgescu**, Rational Choice and Revealed Preference: A Fuzzy Approach
61. **Ping Yan**, Limit Cycles for Generalized Liénard-Type and Lotka-Volterra Systems
62. **Joonas Lehtinen**, Coding of Wavelet-Transformed Images
63. **Tommi Meskanen,** On the NTRU Cryptosystem
64. **Saeed Salehi**, Varieties of Tree Languages
65. **Jukka Arvo**, Efficient Algorithms for Hardware-Accelerated Shadow Computation
66. **Mika Hirvikorpi**, On the Tactical Level Production Planning in Flexible Manufacturing Systems
67. **Adrian Costea**, Computational Intelligence Methods for Quantitative Data Mining
68. **Cristina Seceleanu**, A Methodology for Constructing Correct Reactive Systems
69. **Luigia Petre**, Modeling with Action Systems
70. **Lu Yan**, Systematic Design of Ubiquitous Systems
71. **Mehran Gomari**, On the Generalization Ability of Bayesian Neural Networks
72. **Ville Harkke**, Knowledge Freedom for Medical Professionals – An Evaluation Study of a Mobile Information System for Physicians in Finland
73. **Marius Cosmin Codrea**, Pattern Analysis of Chlorophyll Fluorescence Signals
74. **Aiying Rong**, Cogeneration Planning Under the Deregulated Power Market and Emissions Trading Scheme
75. **Chihab BenMoussa**, Supporting the Sales Force through Mobile Information and Communication Technologies: Focusing on the Pharmaceutical Sales Force
76. **Jussi Salmi**, Improving Data Analysis in Proteomics
77. **Orieta Celiku**, Mechanized Reasoning for Dually-Nondeterministic and Probabilistic Programs
78. **Kaj-Mikael Björk**, Supply Chain Efficiency with Some Forest Industry Improvements
79. **Viorel Preoteasa**, Program Variables – The Core of Mechanical Reasoning about Imperative Programs
80. **Jonne Poikonen**, Absolute Value Extraction and Order Statistic Filtering for a Mixed-Mode Array Image Processor
81. **Luka Milovanov**, Agile Software Development in an Academic Environment
82. **Francisco Augusto Alcaraz Garcia**, Real Options, Default Risk and Soft Applications
83. **Kai K. Kimppa**, Problems with the Justification of Intellectual Property Rights in Relation to Software and Other Digitally Distributable Media
84. **Dragoş Truşcan**, Model Driven Development of Programmable Architectures
85. **Eugen Czeizler**, The Inverse Neighborhood Problem and Applications of Welch Sets in Automata Theory

86. **Sanna Ranto**, Identifying and Locating-Dominating Codes in Binary Hamming Spaces
87. **Tuomas Hakkarainen**, On the Computation of the Class Numbers of Real Abelian Fields
88. **Elena Czeizler**, Intricacies of Word Equations
89. **Marcus Alanen**, A Metamodeling Framework for Software Engineering
90. **Filip Ginter**, Towards Information Extraction in the Biomedical Domain: Methods and Resources
91. **Jarkko Paavola**, Signature Ensembles and Receiver Structures for Oversaturated Synchronous DS-CDMA Systems
92. **Arho Virkki**, The Human Respiratory System: Modelling, Analysis and Control
93. **Olli Luoma**, Efficient Methods for Storing and Querying XML Data with Relational Databases
94. **Dubravka Ilić**, Formal Reasoning about Dependability in Model-Driven Development
95. **Kim Solin**, Abstract Algebra of Program Refinement
96. **Tomi Westerlund**, Time Aware Modelling and Analysis of Systems-on-Chip
97. **Kalle Saari**, On the Frequency and Periodicity of Infinite Words
98. **Tomi Kärki**, Similarity Relations on Words: Relational Codes and Periods
99. **Markus M. Mäkelä**, Essays on Software Product Development: A Strategic Management Viewpoint
100. **Roope Vehkalahti**, Class Field Theoretic Methods in the Design of Lattice Signal Constellations
101. **Anne-Maria Ernvall-Hytönen**, On Short Exponential Sums Involving Fourier Coefficients of Holomorphic Cusp Forms
102. **Chang Li**, Parallelism and Complexity in Gene Assembly
103. **Tapio Pahikkala**, New Kernel Functions and Learning Methods for Text and Data Mining
104. **Denis Shestakov**, Search Interfaces on the Web: Querying and Characterizing
105. **Sampo Pyysalo**, A Dependency Parsing Approach to Biomedical Text Mining
106. **Anna Sell**, Mobile Digital Calendars in Knowledge Work
107. **Dorina Marghescu**, Evaluating Multidimensional Visualization Techniques in Data Mining Tasks
108. **Tero Säntti**, A Co-Processor Approach for Efficient Java Execution in Embedded Systems
109. **Kari Salonen**, Setup Optimization in High-Mix Surface Mount PCB Assembly
110. **Pontus Boström**, Formal Design and Verification of Systems Using Domain-Specific Languages
111. **Camilla J. Hollanti**, Order-Theoretic Mehtods for Space-Time Coding: Symmetric and Asymmetric Designs
112. **Heidi Himmanen**, On Transmission System Design for Wireless Broadcasting
113. **Sébastien Lafond**, Simulation of Embedded Systems for Energy Consumption Estimation
114. **Evgeni Tsivtsivadze**, Learning Preferences with Kernel-Based Methods
115. **Petri Salmela**, On Commutation and Conjugacy of Rational Languages and the Fixed Point Method
116. **Siamak Taati**, Conservation Laws in Cellular Automata
117. **Vladimir Rogojin**, Gene Assembly in Stichotrichous Ciliates: Elementary Operations, Parallelism and Computation
118. **Alexey Dudkov**, Chip and Signature Interleaving in DS CDMA Systems
119. **Janne Savela**, Role of Selected Spectral Attributes in the Perception of Synthetic Vowels
120. **Kristian Nybom**, Low-Density Parity-Check Codes for Wireless Datacast Networks
121. **Johanna Tuominen**, Formal Power Analysis of Systems-on-Chip
122. **Teijo Lehtonen**, On Fault Tolerance Methods for Networks-on-Chip
123. **Eeva Suvitie**, On Inner Products Involving Holomorphic Cusp Forms and Maass Forms
124. **Linda Mannila**, Teaching Mathematics and Programming – New Approaches with Empirical Evaluation
125. **Hanna Suominen**, Machine Learning and Clinical Text: Supporting Health Information Flow
126. **Tuomo Saarni**, Segmental Durations of Speech
127. **Johannes Eriksson**, Tool-Supported Invariant-Based Programming

128. **Tero Jokela**, Design and Analysis of Forward Error Control Coding and Signaling for Guaranteeing QoS in Wireless Broadcast Systems
129. **Ville Lukkarila**, On Undecidable Dynamical Properties of Reversible One-Dimensional Cellular Automata
130. **Qaisar Ahmad Malik**, Combining Model-Based Testing and Stepwise Formal Development
131. **Mikko-Jussi Laakso**, Promoting Programming Learning: Engagement, Automatic Assessment with Immediate Feedback in Visualizations
132. **Riikka Vuokko**, A Practice Perspective on Organizational Implementation of Information Technology
133. **Jeanette Heidenberg**, Towards Increased Productivity and Quality in Software Development Using Agile, Lean and Collaborative Approaches
134. **Yong Liu**, Solving the Puzzle of Mobile Learning Adoption
135. **Stina Ojala**, Towards an Integrative Information Society: Studies on Individuality in Speech and Sign
136. **Matteo Brunelli**, Some Advances in Mathematical Models for Preference Relations
137. **Ville Junnila**, On Identifying and Locating-Dominating Codes
138. **Andrzej Mizera**, Methods for Construction and Analysis of Computational Models in Systems Biology. Applications to the Modelling of the Heat Shock Response and the Self-Assembly of Intermediate Filaments.
139. **Csaba Ráduly-Baka**, Algorithmic Solutions for Combinatorial Problems in Resource Management of Manufacturing Environments
140. **Jari Kyngäs**, Solving Challenging Real-World Scheduling Problems
141. **Arho Suominen**, Notes on Emerging Technologies
142. **József Mezei**, A Quantitative View on Fuzzy Numbers
143. **Marta Olszewska**, On the Impact of Rigorous Approaches on the Quality of Development
144. **Antti Airola**, Kernel-Based Ranking: Methods for Learning and Performace Estimation
145. **Aleksi Saarela**, Word Equations and Related Topics: Independence, Decidability and Characterizations
146. **Lasse Bergroth**, Kahden merkkijonon pisimmän yhteisen alijonon ongelma ja sen ratkaiseminen
147. **Thomas Canhao Xu**, Hardware/Software Co-Design for Multicore Architectures
148. **Tuomas Mäkilä**, Software Development Process Modeling – Developers Perspective to Contemporary Modeling Techniques
149. **Shahrokh Nikou**, Opening the Black-Box of IT Artifacts: Looking into Mobile Service Characteristics and Individual Perception
150. **Alessandro Buoni**, Fraud Detection in the Banking Sector: A Multi-Agent Approach
151. **Mats Neovius**, Trustworthy Context Dependency in Ubiquitous Systems
152. **Fredrik Degerlund**, Scheduling of Guarded Command Based Models
153. **Amir-Mohammad Rahmani-Sane**, Exploration and Design of Power-Efficient Networked Many-Core Systems
154. **Ville Rantala**, On Dynamic Monitoring Methods for Networks-on-Chip
155. **Mikko Pelto**, On Identifying and Locating-Dominating Codes in the Infinite King Grid
156. **Anton Tarasyuk**, Formal Development and Quantitative Verification of Dependable Systems
157. **Muhammad Mohsin Saleemi**, Towards Combining Interactive Mobile TV and Smart Spaces: Architectures, Tools and Application Development
158. **Tommi J. M. Lehtinen**, Numbers and Languages
159. **Peter Sarlin**, Mapping Financial Stability
160. **Alexander Wei Yin**, On Energy Efficient Computing Platforms
161. **Mikołaj Olszewski**, Scaling Up Stepwise Feature Introduction to Construction of Large Software Systems
162. **Maryam Kamali**, Reusable Formal Architectures for Networked Systems
163. **Zhiyuan Yao**, Visual Customer Segmentation and Behavior Analysis – A SOM-Based Approach
164. **Timo Jolivet**, Combinatorics of Pisot Substitutions
165. **Rajeev Kumar Kanth**, Analysis and Life Cycle Assessment of Printed Antennas for Sustainable Wireless Systems
166. **Khalid Latif**, Design Space Exploration for MPSoC Architectures

167. **Bo Yang**, Towards Optimal Application Mapping for Energy-Efficient Many-Core Platforms
168. **Ali Hanzala Khan**, Consistency of UML Based Designs Using Ontology Reasoners
169. **Sonja Leskinen**, m-Equine: IS Support for the Horse Industry
170. **Fareed Ahmed Jokhio**, Video Transcoding in a Distributed Cloud Computing Environment
171. **Moazzam Fareed Niazi**, A Model-Based Development and Verification Framework for Distributed System-on-Chip Architecture
172. **Mari Huova**, Combinatorics on Words: New Aspects on Avoidability, Defect Effect, Equations and Palindromes
173. **Ville Timonen**, Scalable Algorithms for Height Field Illumination
174. **Henri Korvela**, Virtual Communities – A Virtual Treasure Trove for End-User Developers
175. **Kameswar Rao Vaddina**, Thermal-Aware Networked Many-Core Systems
176. **Janne Lahtiranta**, New and Emerging Challenges of the ICT-Mediated Health and Well-Being Services
177. **Irum Rauf**, Design and Validation of Stateful Composite RESTful Web Services
178. **Jari Björne**, Biomedical Event Extraction with Machine Learning
179. **Katri Haverinen**, Natural Language Processing Resources for Finnish: Corpus Development in the General and Clinical Domains
180. **Ville Salo**, Subshifts with Simple Cellular Automata
181. **Johan Ersfolk**, Scheduling Dynamic Dataflow Graphs
182. **Hongyan Liu**, On Advancing Business Intelligence in the Electricity Retail Market

Turku Centre *for* Computer Science

Joukahaisenkatu 3-5 B, 20520 Turku, Finland | www. tucs.fi

University of Turku
Faculty of Mathematics and Natural Sciences
- Department of Information Technology
- Department of Mathematics and Statistics

Turku School of Economics
- Institute of Information Systems Science

Åbo Akademi University
Division for Natural Sciences and Technology
- Department of Information Technologies

ISBN 978-952-12-3093-6
ISSN 1239-1883

Hongyan Liu

On Advancing Business Intelligence in the Electricity Retail Market