# A decision support system for the environmental impact of e-business

**Lan Yi**

Geoenvironmental Research Centre

Cardiff School of Engineering

Cardiff University

Thesis submitted in candidature for the degree of Doctor of Philosophy at Cardiff University

January 2007

# *Acknowledgements*

I would like to sincerely thank my supervisor Prof Hywel Thomas for his continued support and invaluable guidance throughout the duration of the research. I thank him profusely for the opportunity he has offered me in realising this milestone.

For his selfless and valuable technical input I would also like to thank Dr Michael Packianather. Without your inspiration none of this would have happened.

I would like to thank Prof Raymond N Yong for his valuable contribution at the initial stage of this project. His guidance and support as the former supervisor of the author is gratefully acknowledged.

I thank also my family for their continued moral support and for unceasingly bearing the brunt of my absence from China.

To Adil, you have been an endless source of encouragement and motivation. Your support during this period has meant a lot to me in achieving this goal.

Lastly but not the least, to all my colleagues at the Geoenvironmental Research Centre I owe my gratitude for their valuable help.

***Dedicated***

***To***

*My beloved mother Fu Xiangmei and my beloved father Yi Fazhen*
*Adil Muzafar Ganai*

# Summary

With less than half a century's development, e-business and the Information and Communication Technologies it relies on, have been growing rapidly. With an even shorter history than the technology itself, the study of its impact on the environment and sustainable development in general, is still in its infancy. A review of past literature has revealed that the problem is complex. Both negative and positive impacts have been identified. Traditional systematic approaches have been found to be insufficient for this research topic. To explore the relationship further, a new methodology is proposed in this thesis. In particular the main objective of this PhD study is to demonstrate and develop an Expert Decision Support System at the meso level, to simulate the relationship between e-business and the environment.

In pursuit of this aim, results are presented of two surveys that were conducted to collect data and build a knowledge base. Analysis of the data using various techniques was considered, based on data mining technologies and Fuzzy Logic. The development of the Expert Decision Support System is then discussed, adopting a two-way simulation approach. The forward chain of the system is developed based on Decision Support System technology, with the heart of the system built on Neural Networks. Calculation, estimation and prediction of environmental indicator values based e-business indicators are conducted in this part. The backward chain is based on Expert System technology, where conditions and rules are presented to reach certain pre-defined environmental targets. An individual company should then be able to use this system within a certain industry, for example, to simulate its environmental performance by adopting or limiting Information and Communication technologies.

A demonstration of how the system can be used and operated on various occasions for different purposes is presented, based on four application scenarios: predictions, simulations, comparisons and solutions. It is claimed that the results from the Expert Decision Support System, which ideally should be integrated into a company's financial system and other information management systems, will provide important information that could be incorporated into a company's strategic plans, action plans and technological reformation. The research presents a pilot study which tries to not only build a quantitative model but also to construct a decision support system to simulate this relationship in the real world. It is claimed that the work both extends research methodologies in this field and endows traditional Neural Network applications with new meanings and challenges.

# Contents

# Chapter 1

# Introduction

Despite the boom and burst of the dot.com phenomenon, e-business is arousing more interest today than ever before. The whole world seems to be racing to promote business, politics and entertainment online, to as wide an audience as possible. Supported by more and more advanced Information and Communication Technologies (ICT), e-business and ICT are penetrating into every corner of our society and revolutionising the way we live. The growth seems inevitable and has not shown any sign of slowing (Yi and Thomas 2006).

At the same time that this technological revolution is taking place, a green revolution led by the "Sustainable Development" concept is also taking place. The first international Earth Summit was held in 1992, when more than 100 countries met in Rio de Janeiro, Brazil to address the urgent problems of environmental protection and social and economic development. Since that time, sustainable development has been on the agenda of most countries, aiming to achieve economic growth and social development without degrading the environment. However, despite the enthusiasm of environmentalists and governments, many industries are still reluctant to implement the sustainable development concept in their daily business, because environmental responsibility possibly means compromising financial profit (Yi and Thomas 2006).

Seen as a breakthrough by some studies, ecological modernisation theory indicates a possibility of resolving the conflict of economic development and environmental protection by technology. Proponents of ecological modernization assert that it is possible and desirable to both develop economically and socially and at the same time conserve the environment. It is suggested that this can be achieved primarily through technological advances that help to reduce the consumption of resources via increasing efficiency, i.e. pollution prevention, waste reduction. Unlike traditional curative and repair options, the theory focuses on changing social practices and institutional developments associated with environmental deterioration and reform, incorporating environmental considerations at the design stage of technological and organisational innovations (Mol, 1999).

Therefore, could ICT/e-business be one of the technologies to resolve the internal conflicts of sustainable development? Is this technological and business revolution facilitating or decelerating the green revolution? On the other hand, how green and sustainable is this technology itself? Anything growing as fast as e-business/ICT will not be economic, social or environmental indifferent. It is therefore likely to interact with sustainable development.

These questions were the interesting points that triggered this research and motivated the author to study this area. This chapter will explain the research background, which is the real world problem, the scientific/research context and challenges, how the pathway of this study was shaped and the objectives of this PhD study. Finally an overview of this PhD thesis will be given.

## *1.1 Research Background*

Online sales in the UK almost doubled in 2003 to £39.5 billion, up from £19 billion in 2002, according to a report by the Office for National Statistics. Internet sales in the USA reached $21.4 billion in the fourth quarter of 2004, an increase of 22.3% from the same period during 2003 (IMRG 2004). The rapid growth of e-business and ICT can be observed by many facts and it has been affecting the three pillars of sustainable development profoundly.

From the macro-economic point of view, e-business and ICT are making a significant contribution to the boost of the whole economy and productivity. For example,

- Figures published by the US Department of Commerce suggest that in the years 1994–1999 the ICT industries – key enablers of e-business – were responsible for more than 30% of the US real economic growth (Norton 2001)

- Adoption of e-business in the UK, France and Germany has resulted in current cumulative cost savings of €9 billion (US $8.3 billion). These organisations estimate Internet business solutions have also helped to enhance their revenues by €86.4 billion (US $79 billion) to date (Net-Impact-Study 2002)

Also, ICT and e-business are changing the social aspects and improving the quality of our life too. For instance,

- The British government aimed to provide all government services online by 2005 – transacting and procuring as far as possible, and promised over 99% of the population will have broadband services available (Cabinet-Office 2005).

- In South Korea, the world's most advanced broadband country, 70% of homes have broadband accounts, and at peak times just about all of those homes are online. Nearly two-thirds of Korean mobile phone users have shifted to 3G phones, which led to the victory of Roh Moo Hyun, a candidate with a little support from either the mainstream media or the nation's conglomerates, in the presidential election in 2003 (Fulford 2003).

On the environmental side, the impacts of e-business/ICT are much more complicated and mixed. Just to name a few,

- According to DEFRA, powering 86 PCs for a year uses 23,000 kWh of electricity and can generate 10 tonnes of $CO_2$ equivalent emissions (DEFRA 2000a). As a more serious consequence, Green House Gas (GHG) emissions resulted from all these energy consumptions contribute to climate change and global warming.

- It is estimated that by 2010, e-materialisation of paper, construction and other activities could reduce the US industrial energy and GHG emissions by more than 1.5% (Romm et al. 1999).

- Internet economy could render unnecessary as much as 3 billion square feet of buildings – some 5% of the US commercial floor space – which would likely save a considerable amount of construction-related energy too (Romm et al. 1999).

- BT's flexi-working programme 'Options 2000' found teleworking could save 424,000 miles a week of car travel and 190,000 miles a week of rail travel. Teleworking could reduce congestion and commuting traffic and ease the country's transport demands (BT 2003).

- Waste Electrical and Electronic Equipments (WEEE) now constitutes one of the fastest growing waste streams in the EU, with current estimates indicating a rate of growth three times that of municipal waste. In 1998, 6 million tonnes of WEEE were generated (4% of the municipal waste stream). This is likely to have doubled by 2010 (European-Commission 2000).

As one can see, e-business and ICT technologies have been changing and affecting our economical progress, social development and environment protection. Although both positive and negative aspects have been identified, the full list of impacts hasn't been recognised yet, nor is the magnitude of these effects. Regarding whether one side of effects would outweigh the other, no conclusion could be found just yet.

## 1.2 The Scientific/Research Context

The above section has presented the problem in real terms. So what is the scientific problem and challenge for the research community, including scope, angle, methodology and aim of research?

Looking back at the history of communication and technological revolution, electricity took almost a century before it became commonplace. Radio and television was faster. The Internet has existed for 35 years. The World Wide Web appeared only a decade ago. However their ubiquity already makes one wonder how it was possible to manage in the past without accessing such technologies. However, ICT experts, especially those working at the cutting edge, typically lack awareness of the problems that professionals working in areas of human development encounter, and vice versa.

Nevertheless, from past experience, the rapid diffusion of the Internet and new communications technologies suggest that innovations from ICT for sustainable development, economically, socially and environmentally, can also be faster than the progression shown by earlier technologies (Tongia et al. 2004).

Researchers from different backgrounds have presented both environment friendly and environment damaging cases of e-business. In the author's opinion, most of the work conducted to date has aimed to map out the important linkages between e-business/ICTs and the environment. Some research studies have been completed, ranging from relatively small-scale case studies such as "A case-study concerning online sales of personal computers in The Netherlands" (Reijnders and Hoogeveen 2001), to a very large pan-European project such as "Digital Europe" (Digital-EU 2006). The range of the relationship also spans from strictly e-commerce to all encompassing digital technologies and electronic business platforms, and from a sub-sector of the environment to general economic, social and environmental sustainable development. The results of these studies have started to build the knowledge base for this problem.

The currently dominant approach is either a micro-level case study approach or a macro-level statistical approach, both only presenting comparison results or overall outcomes of the impacts. The current macro statistics approach provides a general picture regarding whether ICT/e-business is "good" or "bad" for the economy, society and the environment, but it has been unable to answer what action is needed to promote the "good" and avoid the "bad" by each component (individual person, company, industry and sector) in the society. On the other hand, micro level case studies offer limited contributions due to their restriction to certain specific products, companies, geographic areas etc.

The difficulty, from a macro point of view, in relation to whether e-commerce is affecting a country's environment positively or negatively, is related to demography, economy, policy, geography, culture and many other complicated factors. From a micro point of view, how a company/community's e-commerce adoption and operation affects its environmental performance, is linked to many unidentified, unpredictable or un-measurable factors, due to both unawareness and technical problems. As Abukhader et al. stated, "there is a general agreement that it is highly difficult, if not impossible, to state if the damaging effects of e-commerce on the environment can weigh over the advantageous effects or the contrary." (Abukhader and Jönson 2003).

In the author's opinion, industry or sector based meso-level studies have the potential to not only present the good/bad statistics within an industry/sector, but also offer a solution of how to change what, to achieve certain goal/objectives in the industry/sector. Meso level studies don't exist in this research field at the moment, but they are believed to be able to fill the gap and contribute to our understanding and control of the problem significantly.

In terms of methodologies, qualitative assessments of the positive and negative impacts have been carried out. The challenge of any research is not to just recognise the problem, but to know what can be done, how it can be done, and to choose certain solutions. It is not enough to know ICT/e-business has been changing our daily life, economy, transport, air, water, forests etc. Also it is not even to understand how it is changing everything. Ultimately an approach which can influence the behaviour, say of a company, for example, is needed.

Also a few studies have attempted to develop quantitative indicators for monitoring and measuring progress towards sustainable development (Heinonen et al. 2001). Furthermore, a few models attempt to simulate certain specific parts of the relationship (e.g. (Matthews et al. 2002), (Macauley et al. 2003)). However, these models were developed based on a number of assumptions and have not yet been fully validated for major applications. In the author's opinion these are a good start but still need further consideration. Also a more predictive and empirical model, which can be applied within a sector of society, should be more beneficial in the long term. Such an approach should help simulate potential impacts resulting from changes of indicators, so that positive effects can be promoted and negative ones alleviated. This PhD study is directed towards that goal.

## 1.3 Research Path

The previous section has summarised briefly the research context and challenges. While reviewing other people's work, contributions were acknowledged and limitations were realised. Also a few "inspirations" and advancements have been proposed. So how did these initial ideas help to form the research path for this study?

Recognising that indicators are effectively a tool to examine the relationship between ICT/e-business and the environment, the possibility of assessing indicators in a more generic manner presents itself. The work presented in this study pursues this aim.

In assessing a possible solution to the above problem, it was realised that traditional methods and a systematic approach had an emphasis on methodologies for impact studies, which seem to be insufficient for this research topic. Therefore in the author's opinion, it is time to think outside of the box, so development and demonstration of a new methodology to study the problem was important.

To achieve this aim, the development of an Expert Decision Support System (EDSS) to quantify and simulate the relationship appeared to be potentially useful. If such an approach could be achieved, an individual company would, for example, be able to use this EDSS within a certain industry, to simulate its environmental performance by adopting or limiting ICT technologies. This could include estimation of $CO_2$ emission to stay within the carbon quota and/or comparisons of a company's current environmental performance with similar companies in the industry.

The prerequisite for the development of an EDSS - a knowledge based system, is data. However, data in this field is rather scarce. It was clear from the outset that data collection would form a part of the research. In the author's opinion, it is very critical for academics and governments to standardise the indicators and methodologies and provide best practice for the industries in the long term. In order to build a knowledge base for this research topic, the author conducted two surveys to collect data. Thus the survey and data collection in this study are considered to be a contribution in this area.

Furthermore, in order to build an EDSS in this area, the use of knowledge based techniques such as Artificial Neural Networks, Fuzzy Logic, Decision Support and Expert Systems were investigated.

A neural network is a massively parallel-distributed processor made up of simple processing units, which has a natural propensity for storing experiential knowledge and making it available for use. It resembles the brain in two respects: (i) Knowledge is acquired by the network from its environment through a learning process; (ii) Interneuron connection strengths, know as synaptic weights, are used to store the acquired knowledge (Haykin 1999). It consists of interconnected processing elements called neurons that work together to produce an output function. Neural networks are trainable systems that can "learn" to solve complex problems from a set of examples and generalise the "acquired knowledge" to solve unforeseen problems. So neural networks were proposed in this study, to simulate the relationship between e-business and the environment. With e-business indicators as inputs and environmental indicators as outputs, the networks were trained with survey data and used to predict on unseen situations.

Some data collected from the surveys were typical qualitative descriptions such as "the company's paper consumption has decreased a lot" since it started using ICT. However the term "decreased a lot" cannot be understood by a computer system. Therefore fuzzy logic was introduced in this study to convert linguistic variables to a quantitative number that a computer system can recognise and operate. Fuzzy logic is derived from fuzzy set theory and an extension of Boolean logic dealing with the concept of partial truth where variables can have degrees of truth or falsehood (MathWorks 1995-2005). One of the most commonly used examples of fuzzy set is a set of tall people. If the set of tall people is given the well-defined (crisp) boundary of a classical set, we might say all people taller than 6 feet are officially considered tall. So a person who is 5.999 feet is short while another one with 6.001 feet is tall. It is obvious unreasonable to call one person short and another tall when they differ in height by the width of a hair. In this case, a fuzzy set is more practical. This is a set without a crisp, clearly defined boundary and can contain elements with only a partial degree of membership. So if one asks the question "Is X a member of set A", the answer might be yes, no, or any one of a thousand intermediate values in between.

Fuzzy logic is all about the relative importance of precision. Fuzzy truth represents membership in vaguely defined sets, not likelihood of some event or condition as probabilities. This study attempted to use this fuzzy set and fuzzy reasoning theory for quantifying indicators and reducing data dimension.

The concepts involved in Decision Support Systems were first articulated in the early 1970s by Michael S. Scott Morton (Scott Morton 1971) under the term 'management decision systems' (Turban 1988). Due to its relatively short history, the concept varies greatly: in the narrow sense, it can be defined as "an interactive, flexible, and adaptable computer-based information system, especially developed for supporting the solution of non-structured management problems for improved decision making. It utilizes data, provides an easy-to-use interface, and allows for the decision maker's own insights" (Turban 1995). Adopting a broader definition yields "a computer-based system that aids the process of decision making" (Finlay 1994). In this study, a DSS is defined as an interactive computer system designed to assist the decision-making process with the use of certain analytic methodologies or models. In the case of this study, the proposed DSS consists of a user interface, databases, a data-mining engine and models based on Fuzzy Logic and Neural Networks.

Expert systems are computerised advisory programs that attempt to imitate or substitute the reasoning processes and knowledge of experts in solving specific types of problems. Expert System is a branch of applied Artificial Intelligence and was developed by the AI community in the mid-1960s (Turban 1988). Typical expert system technologies include rule induction, an area of machine learning in which formal rules are extracted from a set of observations. Rule induction was applied in this study to analyse the pattern and trend in a group – what's in common between companies who produce similar environmental performance results. Expert systems are generally designed differently from conventional programs because the problems usually have no algorithmic solution and rely on inferences to achieve a reasonable solution. The knowledge of multiple experts can be made available to work simultaneously and continuously on a problem at any time of day or night (Giarratano and Riley 1998).

This study investigates the possible development of a particular type of Expert Decision Support System (EDSS) – a hybrid system with two-way simulations, containing a so-called "forward chain" and a "backward chain". The forward chain of the EDSS would be based on Decision Support System technology, with the heart of the system built on the implementation of data mining, Fuzzy Logic and Neural Networks. This part of the system would deal with cause-effect (cause: e-business indicators; effect – environmental indicators) calculation, estimation and prediction based upon databases and models. The backward chain of the EDSS would present conditions and rules to be met in order to reach certain pre-defined targets (conditions: e-business indicator range; targets: environmental indicator goals). This part would be based on Expert System technology, with the inference engine built on rule induction and knowledge discovery.

It is suggested that introducing and integrating various knowledge discovery techniques to build a knowledge based Expert Decision Support System for this research topic would be innovative and novel. It is claimed that the work would be the first of its kind and would extend research in this field. It is also suggested that the research would endow traditional Neural Network applications with new meanings and challenges.

## 1.4 Objective and Tasks

The main objective of this PhD study is therefore to demonstrate and develop an Expert Decision Support System, based on Fuzzy Logic and Neural Network technologies, at the meso level, to help a company/organisation analyse how e-business/ICT affects its environmental performance.

This will be achieved through the following tasks:

i) An investigation of the state of the art identifying the relationship between e-business/ICT and the environment, and current methodologies, tools and models to analyse the relationship.

ii) The selection of an industry and the collection of data for relevant indicators representing the relationship.

iii) A qualitative and quantitative analysis of the data. Data mining and knowledge discovery will be employed to find trends and relationships in a data repository and hence build the knowledge base in this research field.

iv) The development of quantitative models to analyse the cause-effect relationship using Fuzzy Logic and Neural Networks, and target-conditions relationships using rule induction.

v) The development of the Expert Decision Support System and the computer program/application encoding the models.

## *1.5 Thesis Overview*

A brief description of each chapter is presented below.

Chapter 2 presents a review of journal papers, theses, projects, project reports, conference and symposiums, and websites, on how various aspects of the environment, such as energy consumption, transport, pollution, waste, material efficiency etc, being affected by various e-business/ICT elements such as teleworking, e-commerce, tele-conferencing etc. A review of methodologies that have been applied or proposed by previous studies is also presented. As research in this field is still in its infancy, some unconventional resources besides the peer-reviewed research are included.

Chapter 3 provides details of data collection and preparation at the initial stage of the project. Details include: why the retail sector and the environmental sector are chosen for survey; the design of the questionnaires; the process of the surveys; the storage of the answers returned; the analysis and processing of the data collected; the qualitative analysis and assessment of the results and issues to be taken into account in evaluating the results. The survey and data collection in this study are considered to be a contribution in this area.

Chapter 4 explains in detail how the data is further processed, mined and modelled. This part of the work forms the engine of this study's Expert Decision Support System. The methodologies adopted include general data mining techniques, Fuzzy Logic and Neural Network. Their concepts, major components and principles are introduced. Also the reasons why these approaches are chosen and how these methodologies fit into this study are discussed. The software choices for the application are presented. An innovative modular neural network approach, based on correlation coefficients between factors, is proposed, tested and found to be feasible and appropriate for this study. Finally, three modular networks are selected, defined, constructed, trained, tested and validated.

Chapter 5 presents the development and implementation of the Expert Decision Support System. A comparison between Decision Support Systems and Expert Systems is made to explain why an expert system is a complement to the DSS and why this hybrid system is required. The forward chain of the EDSS is based on Decision Support System technology, dealing with cause-effect calculation, estimation and prediction based upon databases and models. The backward chain of the EDSS is based on Expert System technology, presenting conditions and rules to be met in order to reach certain pre-defined targets. A Graphical User Interface is built. This step is undertaken to ensure that understanding of background technologies, including Neural Networks, DSS etc, would not be necessary to operate the system.

Chapter 6 initially presents further validation of the EDSS. Three stages of validation and verification are presented as follows: i) checks on the accuracy of the neural network engine, ii) checks on the rules induced and iii) checks that the interface exchanges user inputs and advice properly with the background engine. The latter part of the chapter presents four application scenarios: predictions, simulations, comparisons and solutions. Screenshots are provided to demonstrate how an end user or a company/organisation can use and operate this expert decision support system for various purposes. These include, for example, estimation of $CO_2$ emission to stay within the carbon quota and comparisons of a company's current environmental performance with the average in the industry.

Chapter 7 draws together the major conclusions of this PhD study. Further development and future work are also suggested.

# Chapter 2

# Literature Review

## *2.1 Introduction*

This chapter will present the results of a literature review carried out by the author, focusing on how e-business and related ICT/digital technologies affect the environment.

It is recognised that conventionally a PhD literature review focuses on academic papers that have been peer reviewed. However, in this study, where the terms "e-business", "e-commerce" and "World Wide Web" were only officially established in the 1990s, research is still in its infancy. With only around ten years' history, research in this area is also limited. Based on the fact that figures/facts/analyses quoted by many papers actually come originally from non-academic sources such as project reports and websites, the author believes that reviewing academic papers alone would exclude some important work. So other forms of research output – projects, reports and Internet websites are also included.

The structure of this chapter is organised as following:

Section 2.1 – introduction of the chapter

Section 2.2 – review of journal papers, special issue of journals and thesis that have been peer reviewed. This part of work covers emphases on how various aspects of the environment, such as energy consumption, transport, pollution, waste, material efficiency etc, being affected by various e-business/ICT elements such as teleworking, e-commerce, tele-conferencing etc.

Section 2.3 – review of five projects and corresponding project reports, which are relatively significant breakthrough regarding this research topic from various countries

Section 2.4 – review of methodologies that have been applied or proposed by previous studies

Section 2.5 – review of conference and symposia, within which some important research themes were brought up to academics and some important conference papers with innovative approaches were produced

Section 2.6 – review of websites, some of which provide thorough information resources in this research field and the most substantial work in the author's opinion.

Section 2.7 – overall conclusions of the literature reviewed

Section 2.8 – references of this chapter

## *2.2 Journal papers and thesis*

### 2.2.1 Journal papers

One of the earliest journal papers published in this field is by Jokinen et al, published in 1998 and entitled "The environment in an 'information society': A transition stage towards more sustainable development?". Jokinen et al. examined the relationship between the information society and sustainable development, on theoretical and conceptual levels, stating that information technologies could reduce the stress on the environment by dematerialisation of products and immaterialisation of consumption. They also stated that, on the other hand, positive environmental effects might be overcome by the 'rebound effect' caused by excessive economic growth (Jokinen et al. 1998).

The conclusions made by Jokinen et al were as follows:

i) The relationship was complex and contradictory, and no simple generalisations about the role of informational technologies in environmental policies should be made.
ii) The subject required additional research at the empirical level. Further study was needed, for instance, in order to develop valid indicators that could successfully unite the positive factors of the information society and of sustainable development.
iii) Future projects could also investigate the kind of values and environmental impacts implicit in the authoritative political programmes of the information society.
iv) Finally and most importantly, there should be continuous questioning in the future of how the elements of these basic societal goals should and could be adapted to each other.

Though no definitive conclusions were made, this work still remains of importance in the author's opinion since it raised the issue/problem at such an early stage.

In the same year Nevin Cohen published "Greening the Internet: ten ways e-commerce could affect the environment". He looked at the phenomenal growth of Internet and e-commerce, and stated that "the future of the environment may be profoundly affected for better and for worse" (Cohen 1998). The ten major impacts that were summarised by Cohen are quoted in-order as following:

i) "'Mass Customisation for Eco-Efficiency" ("Just-in-Time" for decreasing warehouse space; "Just Enough" for avoiding over-production; and "Just for You" for reducing materials consumption)
ii) "Marketing by Pixels Instead of Packages"
iii) "Dematerialisation of Products" (from CDs to MP3s, from snapshots to JPEGs, from cheques to clicks)
iv) "The De-Malling of America" (replacement of physical shopping mall with e-commerce)
v) "Let Your Modem Do the Driving" (Internet shopping replacing travelling)
vi) "Closing the Loop On-Line" (cost-effective IT system to reuse packaging and intelligent route design for deliveries)
vii) "GreenBot.com" (green consumerism promoted by the transparent e-commerce ecological comparison)

viii) "Materials Reuse through On-Line Auctions" (reuse second hand goods through online auctions)

ix) "Adding Information to Products for Environmental Efficiency" (after sales information and services to extend product lifecycle)

x) "Global E-commerce" (reduced trade barriers, enhanced communication, and increased technology transfer have led to increase in global commerce, which could have enormous environmental impacts, both positive and negative).

This is one of the earliest works to look at the topic from a theoretical point of view. It is slightly narrower than the previous paper and more focused on e-commerce and environmental sustainability. Its importance again lies with the fact that it provided a starting point for work in this field and raised awareness of the issues involved.

In 2000, Roome and Park gave an introduction in a special issue of Greener Management International (GMI), which attempted to provide a broad context within which a debate about environmental sustainability of technologies in general, and ICCE (Information, communication, computing and electronic) technologies in particular, could be framed. Roome and Park suggested that ICCE technologies should be assessed within the context of their socio-technical systems and actor networks but ought to be considered differently to the other innovations such as automobile innovations, for example, because the technologies themselves were still evolving and the shape of the socio-technical systems that develop around ICCE or the nature of the actor networks were not yet known. Roome and Park believed that ICCE technologies were likely to become more important as they become increasingly ubiquitous over the next 50 years, and ICCE have brought all the dimensions of human systems – economic, financial, cultural, social, political and ideological – into much deeper and more complex relationships with environmental systems.

Roome and Park stated that the other writers of the papers in this special GMI issue could all probably conclude at the moment ICCE technologies would have the potential for both a good and bad outcome, and important implications for sustainability. However they also expressed concerns regarding the companies that were developing ICCE technologies and applications systems, that it was still far from clear the extent to which the issue of sustainable development was being considered (Roome and Park 2000). The unidentified issues were expected to be explored in an edited book (Park and Roome 2002), which at that time had not been published.

In 2001, Reijinders and Hoogeveen published a paper entitled "Energy effects associated with e-commerce: A case-study concerning online sales of personal computers in The Netherlands". In this paper they addressed the energy implications of e-commerce at the micro level, by quantifying transport related energy savings in the case of a Dutch online computer reseller and assessing the extra energy expenditure associated with the increased buying power of online buyers.

It was found that 'energy use' per article sold by the online computer reseller was lower than that sold by the traditional shop. On the other hand Reijinders and Hoogeveen showed that there were scenarios that led to an overall increase in energy use if indirect effects such as increased consumer buying power were taken into account. They however pointed out two main limitations with regard to the foregoing analysis related to PC distribution. First, non-distribution related direct effects of e-commerce were not quantified, as no benchmark data was available. Second, they felt that energy intensities were only a partial reflection of the overall environmental impact of PC usage (Reijnders and Hoogeveen 2001).

Heinonen et al presented a concept which they termed Information Society Assessment (ISA) in a paper published in 2001, entitled "The ecological transparency of the information society". The Fifth Framework Programme of the European Union had largely replaced the terms 'information and communication technologies' and 'telematics' with the broader term: Information Society Technologies (IST). Heinonen et al thus proposed ISA as an assessment tool to examine approaches and activities, to evaluate the implications of IST.

They developed a few scenarios, criteria and indicators as tools for identifying various environmental impacts inherent in the information society. They also suggested that differences among countries could be considerable, so they felt that the establishment of reliable and internationally comparable data required monitoring, analysis, treatment and checking on a continuous basis. Heinonen et al also made what they considered to be a critical point, i.e. that indicators were only a tool for evaluation, linking the indicators more closely to the established goals and commitments was the ultimate challenge (Heinonen et al. 2001).

Miller and Wilsdon published a paper entitled "Digital Futures — An Agenda for a Sustainable Digital Economy" in 2001 (Miller and Wilsdon 2001). This paper was one of the major products of the Digital Futures project, one of the most significant projects in this research field. In this paper, findings were summarised in the form of ten headline principles – what they could call their ten 'dot-commandments'. The ten headline principles were:

i) "Beyond the hype, there's hope" (E-commerce created new opportunities for environmental and social sustainability)
ii) "The e-economy can access all areas" (The digital revolution could refresh the parts that other revolutions haven't reached, by spreading benefits to all regions of the UK and all sectors of society.)
iii) "Community is alive & clicking"(Online relationship, supported by e-commerce, could add a valuable extra dimension to real world interaction)
iv) "'e' is for environment" (E-commerce could help to cut energy and resource use, and improve environmental productivity)
v) "HTML = heavy traffic made lighter?" (Virtual traffic could replace real traffic. With the right policy framework, e-business could create more efficient logistics and distribution systems.)
vi) "Trust me, I'm a dot-com" (E-commerce was changing the relationship between companies and stakeholder, and could usher in a new era of corporate transparency and accountability.)
vii) "But right now, matter matters more" (Potential environmental gains wouldn't be realised without a concerted effort from government and business to align e-commerce with wider sustainability objectives)

viii) "Smart technology needs smart institutions" (Technology was developing at breakneck speed. Institutionally we were struggling to keep up. Sustainable e-business would depend not just on technological innovation, but also on social and political innovation.)

ix) "We need to join the dots" (Partnership would be key to the creation of a sustainable digital economy. Dot-coms, dot-govs and dot-orgs needed to work together more often and in new ways.)

x) "It's about time" (A year in cyberspace was said to be four months. As the Internet accelerated the pace of life, we should change our attitude to time and long term responsibility.)

The overall project, along with other project reports, will be reviewed together in further details later in this chapter, in the project section.

In a paper "Environmental impacts of the emerging digital economy: The E-for-Environment E-Commerce?" published in 2002, Sui and Rejeski analysed the positive environmental impacts of the emerging digital economy as three Ds: dematerialisation, decarbonisation and demobilisation. They stated that some rebound effects were under-appreciated and attempted to evaluate the combined net effects of the digital economy on the environment, taking both negative and positive environmental impacts into consideration. Sui and Rejeski also stated "The honest answer is that we don't know, and worse, that such impacts may not be knowable in the conventional sense because many aspects of the environmental impacts of digital economy defy quantification."

The paper sought to remind environmental policy makers the "daunting complexities and uncertainties" of the relationship, and that optimising the environmental performance of an economy driven by information and knowledge creation was different from regulating one based largely on the processing of material. Sui and Rejeski felt that many of environmental policy tools were too blunt and reactive to steer technological and social innovation in an information economy in which traditional notions of borders, distance, jurisdiction, and time had been altered in very fundamental ways (Sui and Rejeski 2002).

In the same year, Hesse analysed the impact of e-commerce on freight transport, logistics and physical distribution, regarding both Business-to-Business (B2B) and Business-to-Consumer (B2C) e-commerce. Various distribution channels, namely: truck transport efficiency and frequency, bulk freight carriers, airfreight and customer travel were considered.

Only specific, project based conclusions were made as, at least momentarily, it was considered that a general statement on the net effect of e-commerce on freight transport and the environment was almost impossible to make. This was due to three factors: the broad variety of distribution practices (supply side), the unpredictable behaviour of the users (demand side), and weak empirical data both in terms of selected case studies and the broader context of interrelationships and spatial outcome.

Hesse also stated that almost nothing was known about cumulative causation: how new technologies, after becoming embedded into significant practices, were related to their social and economic environment and what their long term, so-called second order effects were (Hesse 2002).

Also in 2002, Matthews et al compared the environmental and economic performance of traditional retailing and e-commerce logistic networks in the United States and Japan, using book retailing as a case study. The work initiated the task of quantitative systems analysis comparing the two logistics systems using i) an Economic Input-Output Life Cycle Assessment (LCA) for the online system and ii) the traditional LCA for the traditional model. Matthews et al discovered that 'crossover points' existed between traditional and e-commerce enabled book retailing so there was no definitive answer as to which method was more energy efficient. Also the quantitative results of the two case studies differed greatly due to the geographical and social differences between the two countries (Matthews et al. 2002).

The results of the two case studies were based on quite a few significant assumptions, such as shipping distances, return rates and shopping purchase allocations in the US case study and energy consumption by computers, travel distances etc for the Japan case study. Therefore in the author's opinion, it is possible that the results might change largely if the assumptions don't stand. However this study made a significant contribution to this field of research because it is one of the earliest attempted to use quantifiable and comparable methods to establish the relationships.

In 2002, Romm published a journal paper entitled "The Internet and the new energy economy", based on work first published in 1999 in a working paper (Romm et al. 1999). These two papers have been the most frequently quoted and referred papers in this field. Romm claimed that recent reductions in energy intensity were connected to the growth in information technology (IT) and the Internet economy. He believed the possible cause of the remarkable change - higher GDP growth with lower energy growth - was IT. As the IT sector itself was less energy intensive and IT also appeared to be increasing efficiency in every other sector of the economy, so the growth in the Internet economy could cut energy intensity, however he stated the conclusion could not be drawn with certainty. The impact of the Internet economy on manufacturing, buildings, and transportation were explored. (Romm 2002)

In 2003 Macauley et al attempted to model the costs and benefits of policies to manage 'e-waste'. They focused on a large component of the electronic waste stream, computer monitors. Environmental concerns in this matter were associated with disposal of the lead embodied in cathode ray tubes (CRTs) used in most monitors. It was found that the benefits of avoiding possible negative health effects associated with CRT disposal appeared far outweighed by the costs, for a wide range of policies.

This study proposed Computer Monitor Policy Simulation, or COMPS, allocated retired CRTs across the different end of life (EOL) discard options based on a cost-minimization algorithm. The algorithm explicitly accounted for the heterogeneity of the costs associated with each of the different options for different classes of CRT users, including both residential and non-residential consumers. A representative sample of two thousand individual members of each category was constructed using a repeated Monte Carlo technique. Implemented using a software package called Analytica, COMPS included four major modules: the CRT and Material Flow Module, the Cost Module, the Environmental Impact Assessment Module and Parameters, and the Index Module.

Macauley et al recognised that CRT glass might break when a computer monitor was placed in a landfill, potentially resulting in lead contamination of groundwater from landfill leachate. The model translated estimates of lead emissions from disposed CRTs to dollar estimates of the health damage caused by air emission. They found a ban on landfilling and incineration of used monitors resulted in an average disposal cost of almost $20 per monitor while a ban on incineration resulted in an average cost of just over $3 per monitor. The benefits of reducing airborne emissions of lead associated with CRT incineration appeared to be small. They concluded that a recycling subsidy was a more cost-effective way to increase monitor recycling than a disposal ban. However, a recycling subsidy reduced storage more than it reduced incineration or landfilling, causing the environmental benefits from reduced incineration to be small (Macauley et al. 2003).

Macauley et al claimed this model was the first of its kind. Without verifying how reliable this model is and though being restricted to the impact of CRT specifically, the author believes their work is still a considerable progress in this research field. The simulation built is considered to be a significant step forward from the other very few quantitative studies in the author's opinion, as it translated the data, which were numerical numbers only, to information that actually makes sense for policy-makers and end-users.

The MIT press journal "Journal of Industrial Ecology" published a special issue – E-commerce, the Internet and the Environment (Volume 6, Number 2) in 2003. A number of articles were presented, examining the problem from different perspectives. To name a few: "E-commerce: Sorting Out the Environmental Consequences"

(Fichter 2003), "Information Technology and U.S. Energy Consumption: Energy Hog, Productivity Tool, or Both?" (John 2003), "Effects of E-Commerce on Greenhouse Gas Emissions: A Case Study of Grocery Home Delivery in Finland" (Siikavirta et al. 2003), "Telecommunications and Travel: The Case for Complementarity" (Mokhtarian 2003).

As the editor Rejeski stated, this special edition of the Journal of Industrial Ecology was an attempt to pull together and critically examine the research on information technology and the environment. The picture that emerged was ambiguous but nonetheless provided a cautionary tale for those who believed IT as a means of environmental salvation or an easy road to sustainable development. He also stated that the research presented in this volume provided no clear answer concerning the effects of information technologies on the environment. These effects, whether positive or negative, were often dependent on a wide range of variables, many of which might be hard to predict or address with traditional environmental policy approaches (Rejeski 2003).

The author believes that Rejeski also raised an important issue on research in this field, that it had become clear that the amount of research being funded to examine the information technology/environment area in no way matched the challenges to our understanding created by the digital revolution. He also stated that probably there was no lack of interested or qualified researchers, but a deficit of research funds needed to be addressed at both national and international levels.

In the author's opinion, this special issue of the journal is a significant milestone in this field. It not only looked at various aspects of the problem itself, but also raised the importance of this research topic to an unprecedented level by an academic journal.

In the Spanish study "The environmental impacts of teleworking: A model of urban analysis and a case study", Pe′rez et al reviewed the literature of the environmental impacts of teleworking and developed a model of the monetary value of the environmental impacts of teleworking for a Spanish city as: "***VIM = ACA + ACT – CCD + ACE***" using four dimensions: annual savings in air pollution (ACA), annual savings in traffic congestion (ACT), environmental costs of urban relocation (CCD), and the net environmental savings of e-commerce (ACE). The ACA was calculated for example, according to:

$$ACA = \left[ \sum_i \left( \{pt * f * \frac{v}{o}\} * em_i * cmc_i \right) \right] * dt \qquad (1)$$

*ACA*: annual savings (euros) from air pollution by teleworking
*pt*: the diffusion rate of teleworking
*f*: the frequency of teleworking
*v*: the use of private vehicles to commute
*o*: the average occupancy rate of automobiles in the area
*emi*: the reduction of air pollution emissions
*cmci*: the marginal cost of pollutant emissions
*dt*: the average number of working days in a year

Pe´rez et al concluded that a significant reduction in the number of kilometres travelled and the emission of pollutants occurred, even after including the rebound effect of teleworker's modifications in travel patterns such as trips around tele-centres. (Pe´rez et al. 2004)

In the author's opinion, although this work reviewed the environmental impacts of teleworking particularly, measured only in monetary term and application is limited to a Spanish city, it is an important advancement in terms of quantitative research in this field, which provided a start for further empirical studies.

Berkhout and Hertin drew distinctions between direct, indirect and structural/behavioural effects of ICT and e-commerce in the paper "De-materialising and re-materialising: digital technologies and the environment" published in 2004. The conclusions were made as: the direct effects of ICT were predominantly negative and stemmed from production, use and disposal of hardware such as computers, screens, network cables etc; the indirect impacts related to the effect of ICTs on production processes, products and distribution systems were expected to be largely positive; and structural/behavioural impacts related to more fundamental processes had both positive and negative outcomes. The net environmental impacts were relying on the balance of de-materialising and re-materialising. Though ICTs did not necessarily lead to a more environmental-sound future, they offered new opportunities to develop more sustainable solutions (Berkhout and Hertin 2004).

Nevertheless, in the author's opinion, the most interesting point of this paper is not that it tried to map out the linkages between ICT and the environment, but the research outlook it proposed. Berkhout and Hertin reckoned much of the literature on the link between ICTs and the environment was highly specific, identifying potential for environmental savings on the basis of isolated case studies. It often involved simple modelling of potential direct effects of the diffusion of a certain technology or technology related phenomenon (if X is substituted with Y, this will lead to Z savings). Although this micro-level research was valuable, it had a number of inherent limitations and stood in the way of general findings. A more systematic analysis of the full complexity of the relationship over the long term was now emerging but it faced a number of methodological challenges. The empirical evidence on which to draw conclusions was still thin.

Therefore they proposed three research approaches: i) macro-level qualitative studies developing and assessing a range of alternative development pathways, and mapping the long-term relationship between environmental sustainability and ICTs; ii) macro-level quantitative studies; and iii) meso-level analysis of sectoral trends and ICT effects that look across a range of sustainability indicators to measure effects.

"Environmental implications of wireless technologies: News delivery and business meetings" is another paper at the case study level with comparison approach published in 2004. In this paper, the environmental effects of two applications of wireless technologies were compared to those of conventional technologies for which they could substitute.

First, reading newspaper content (New York Times daily) on a personal digital assistant (PDA) was compared to the traditional way of reading a newspaper. It showed that the traditional paper format resulted in the release of 32-140 times more $CO_2$, several orders of magnitude more NOx and SOx, and the use of 26-185 times more water depending on whether a newspaper was read by 1 or 2.6 readers and the level of recycled content used in the newsprint. Second, wireless teleconferencing was compared to business travel. The results showed that wireless technologies created lower environmental impacts too. (Toffel and Horvath 2004)

Gay et al developed an environmental input-output life cycle assessment model in 2005 to compare the environmental impacts of a traditional business strategy with an e-commerce strategy for the personal computer industry specifically.

As they claimed, outputs of this model included measures of electricity, natural gas, fuel and packaging expended, retail and warehouse space used, energy expenditures, vehicle emissions and 20 different pollutants. The software the study used was SmartCost, an object-oriented cost modelling system. The simulation ran for 48 months with Dell representing the e-commerce model, and results were presented with tables: transportation mode emissions based on distance travelled, life cycle comparisons (energy consumed and pollution caused) of traditional business and four scenarios of e-commerce with various percentage shipments by air etc. (Gay et al. 2005)

Gay et al made a very positive conclusion that for the personal computer industry, e-commerce was much more environmental beneficial than the traditional counterpart. However, in the author's opinion, the conclusion was made within a specific industry sector, or even with a specific e-commerce 'representative' - Dell, and the comparison made wasn't based on the same geographic area either, which might affect the results. Furthermore, the model was very sensitive to the amount of air freight, distance travelled, size of items being shipped etc. However Gay et al were confident about the potential of applying this model in other industries and examining the global environmental impact of e-commerce for many trades (e.g. online banking and governmental services).

In the same year Wager et al analysed the implications of the disposal and recycling of packing materials containing so-called smart labels in Switzerland, which were very thin and flexible self-adhesive tags with an integrated passive RFID (Radio Frequency IDentification) transponder that comprise one important development track in the framework of ubiquitous or pervasive computing. Wager et al also discussed the results from the perspective of the Precautionary Principle, which aimed at anticipating and minimizing potentially serious or irreversible risks under conditions of uncertainty. The Precautionary Principle was considered being incorporated into many international treaties and pieces of national legislation for environmental protection and sustainable development, but rarely been applied to novel ICT and their potential environmental impacts.

Wager et al argued that a broad application of smart labels bore some risk of dissipating both toxic and valuable substances, and of disrupting established recycling processes. However the risks could be avoided by precautionary measures, mainly concerning the composition and the use of smart labels. They proposed to implement these measures as early as possible in order to avoid irreversible development that was undesirable from the viewpoint of resource management and environmental protection. Although the findings referred to the Swiss waste management system, they were supposed to be transferable to other countries with modern, industrial waste management systems. (Wager et al. 2005)

In 2006 Hilty et al published a paper entitled "The relevance of information and communication technologies for environmental sustainability – A prospective simulation study"- the summary of the major results of the project commissioned by the Institute for Prospective Technological Studies (IPTS) of the European Commission (Hilty et al. 2006).

Hilty et al categorised the impacts of ICT on the environmental sustainability on three levels: i) First-order effects such as increasing electronic waste streams; ii) second-order effects such as improved energy-efficiency of production; iii) third-order effects such as a product-to-service shift in consumption or rebound effects in transport. In the simulation study described in this article, they modelled all known relevant effects on all three levels using a System Dynamics approach in combination with scenario techniques and expert consultations.

However, they felt some limitations of System Dynamics became apparent during the project. They stated that the lack of true modularity made it almost impossible to control the growing complexity of the model. The absence of the concept of a (complex) simulation experiment made it inevitable to mess up the model with control elements and auxiliary equations. Due to these limitations of the approach chosen, the model lost almost all communicative value for discussions, which was contrary to the original intention.

Nevertheless, Hilty et al believed that the project was the first approach to account for the complex causation structure of the environmental effects of ICT in a prospective study; it generated some robust results that paved the way for future research and decision support in this field. They stated that the results should not be interpreted as forecasts of the development of the environmental indicators, because their absolute values in 2020 greatly depend on the three scenarios chosen and on many uncertain model parameters. However, the relative change of the indicators under the influence of ICT, in particular regarding the combined effects of specific ICT application areas, turned out to be quite robust when scenarios and uncertain parameter were varied to test the sensitivity.

Therefore they thought that this project contributed to the general understanding of the environmental impacts of ICT and provided a useful basis for policy-making in the fields of ICT and environment.

The overall IPTS project and the five original interim reports that this paper was based on will be explained further in the project review section 2.3.4.

In conclusion, journal papers in this research area, most substantial ones rather than an exhaustive list, have been reviewed in details in this section. It is obvious that the research is still in its infancy in this field. As Abukhader and JoÈnson summarised, "the literature on this topic (environment and e-commerce) is fertile of pilot studies and propositions, but with very limited support of empirical evidence. Scholars of different backgrounds have studied the impact of e-commerce on the environment differently, in terms of approach, assumptions and demarcations" (Abukhader and Jönson 2003).

As expected, the amount of papers that make a significant difference is very limited and far from finding any definitive conclusions or solutions, and depth of research is also not enough, but on the positive side, significant research interests have definitely been provoked and awareness of issues involved has also been raised not only in academics but also to policy makers.

The papers reviewed above have shaped a good and interesting path for future research from various angels. Among these limited work, controversial evidence and discussions were found, however this is the case all new research topic and pilot studies would face. Researchers who have started looking into this issue already realised that the problem is not and not going to be straightforward, therefore it induced a new research field that is well worth being studied further.

### 2.2.2 Theses

In 1999, Arnfalk published a thesis entitled "Information Technology in Pollution Prevention" in Sweden. This PhD dissertation attempted to evaluate how teleworking and teleconferencing (video-conferencing and audio-conferencing) could be used by an organisation to prevent or reduce its environmental impact. The thesis attempted to address the following questions:

- Why should a company or other organisation use teleworking and teleconferencing to reduce its environmental impacts?
- What potential environmental and economic savings could the use of these applications generate?
- What were the present environmental and economic implications of the use of these applications?

- What were the main barriers and facilitators towards obtaining these savings?
- What should an organisation do to effectively use the applications teleworking and teleconferencing to reduce its environmental impacts?

Within this research, a number of surveys were conducted, designed to evaluate the costs and benefits of teleworking and teleconferencing from two main perspectives: the environmental effects and economic effects. The main factors inhibiting and promoting the use of teleworking and teleconferencing as tools for pollution prevention were identified and divided into personal, organisational, institutional and practical factors.

Based on experiences collected from diverse studies, Arnfalk developed a model for estimating environmental and economic effects from a number of selected impact categories, namely transportation, office space, equipment and communication. The model was applied to a fictitious company, *Virtually Green,* for which two development scenarios were constructed: i) 'Business as Usual' of how organisations currently utilise teleworking and teleconferencing, and ii) 'Collabication' of how the applications could be used with an added mix of new ideas and solutions (Arnfalk 1999).

In the author's opinion, this thesis provided a comprehensive view of how ICT, teleworking and teleconferencing specifically, influenced the environment. The model proposed, however, would be more convincing if tested on a real case rather than an imaginary one. It is still quite an important piece of work at such an early stage of this field of research. With more data and more studies available since then, the methodology has the potential to be improved and verified significantly.

In 2001, Hurst published a thesis entitled "E-commerce and the Environment" and examined the involvement of environmental issues into the B2C (business-to-consumer) e-commerce (Hurst 2001). With 7 British company case studies (The Body Shop, ThinkNatural, NatWest bank, Smile, Thomson Travel Group, WHSmith and Ethical- Junction), the research looked into issues like what e-commerce leaders and environmental managers thought about the environmental implications of using such a platform, whether or not companies integrated existing environmental strategies into their e-commerce business, and the long term benefits of doing so.

Direct environmental impacts of e-commerce were identified in 9 aspects: energy use and Green House Gas (GHG) Effect; waste electronic and electrical equipment (WEEE); paper demand; transport patterns and pollution; land use pattern and infrastructure; carbon trading; demands for raw materials and products; demand for green products; life cycle analyses. Indirect environmental impacts of e-commerce were identified as environmental accountability, awareness and reputation.

Hurst gathered, presented and examined both positive and negative debates and information about each effect of e-commerce identified, from media coverage, published research, company interviews etc.

Due to lack of qualitative evidence, methodological rigor and limitations in the data set, this research was unable to answer some questions it originally set out to solve, which include: whether or not e-commerce would have a net negative or positive affect of environmental sustainability; the extent of consumer demand for environmental information; the business case for integrating environmental policies and performance into e-commerce operations.

In 2003, Abukhader from Lund University (Sweden) produced a doctorate thesis entitled "The Environmental Implications of Electronic Commerce – The Assessment Approach Problem" in the form of a collection of four journal papers in 2003 (Abukhader 2003). The two papers most relevant to this subject are:

PAPER 1: "The Environmental Implications of Electronic Commerce – A Critical Review and Framework for Future Investigations" (Abukhader and Jönson 2003)

PAPER 2: "E-commerce and the Environment: A Gateway to the Renewal of Greening Supply Chains" (Abukhader and Jönson 2004)

In the first paper, literature in the topic of e-commerce and environment from 27 main journals in the period of 1991 – 2002 were surveyed and reviewed, across three fields:

logistics/supply chain management, the environmental science, and e-commerce/business. The research categorised the identified literature review in terms of different view angles of the problem, including e-commerce resulted transport emissions, packaging and warehousing, energy consumption and paper consumption.

Abukhader and JoÈnson suggested looking at the e-commerce impacts in this regard from the Supply Chain Management (SCM) point of view, with the environmental Life Cycle Analysis (LCA) method.

In the author's opinion, this paper presented a fairly thorough picture and overview about the research done of the time and categorised them in a very sensible order. Along with the reference list, this paper provided a base for the background preparation of this PhD research.

The second paper was mainly focused on greening supply chains and further suggested adoption of a two-dimensional model instead of the traditional LCA assessment. The model consisted of the vertical pillar of LCA, and the horizontal line of four main physical processes: production/manufacturing, packaging, transportation, and warehousing. Rather than an innovative method Abukhader and Jönson claimed the model to be, it was essentially still a LCA approach in the author's opinion, just different products/processes were differentiated and separated at difference stage of life cycle. Further examples/case studies would help understand how to implement this model/concept practically.

The main focus of this Licentiate was to study the environmental implications of only e-commerce, rather than other type of information technology. Abukhader evaluated his own work as: rather than perform an assessment of environmental implications, he studied how to approach the assessment, so only the assessment issue itself was discussed and examined. Moreover, Abukhader discussed and proposed a model, only for the assessment of the secondary effects of e-commerce, which was mainly the impact of the use and applications of e-commerce, such as influence over systems of transportation, warehousing, packaging etc, rather than the primary effects (impact of building e-commerce equipment hardware etc) and tertiary effects (rebound effects due to change of human behaviour induced by e-commerce).

In the author's opinion the main contribution of this Licentiate work is located in his effort to contribute in structuring this area and offering ideas, models and thoughts for a next developing research stage. It is often the case that such type of research begins with a group of pilot/short studies as a starting step.

## 2.3 Projects Review

### 2.3.1 Digital Futures, 2001

The Digital Futures project was launched in February 2000 as a year-long inquiry into the social and environmental opportunities of e-commerce and the new economy. Co-coordinated by Forum for the Future, the project was a partnership between 14 corporate partners (amazon.co.uk, AOL, BP, BT, Ericsson, Kingfisher, Nationwide, The Post Office, Royal Bank of Scotland, Royal and Sun Alliance, Smart South West, SUN Microsystems, Unilever, WHSmith), 3 government departments (Office of the e-Envoy, Department of Trade and Industry, Department of the Environment, Transport and the Regions) and 8 think-tanks / research organisations (Demos, Forum for the Future, Green Alliance, Local Futures Group, SPRU, TCPA, UK CEED/University of Bradford, New Economics Foundation, Town & Country Planning Association).

Forum for the futures, one of the eight think-tanks involved, published the "Digital Futures: an agenda for a sustainable digital economy" (Miller and Wilsdon 2001) which has been reviewed in the journal paper section of this chapter.

The major results of the project were presented in the following table, with four scenarios against various indicators. The four scenarios (Figure 2-1) were defined by two dimensions, the horizontal *values* dimension meaning contemporary tastes, beliefs and norms; the vertical *governance* dimension defining the way in which authority and control was exercised in societies – whether local, national or global.

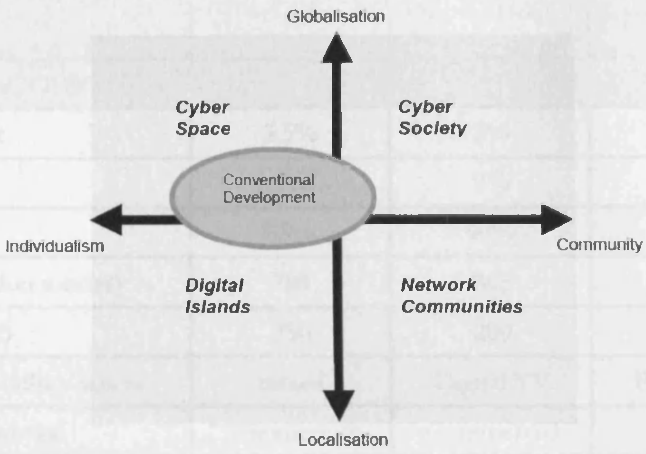



**Figure 2-1 Four contextual e-commerce scenarios**

| Indicator | CyberSpace | DigitalIslands | CyberSociety | Networked Communities |
|---|---|---|---|---|
| SCENARIO CHARACTERISTICS | | | | |
| GDP growth per year | 3.5% | 2% | 2.75% | 1.25% |
| Unemployment | 6% | 9% | 5% | 6% |
| Private car use | 85% | 90% | 78% | 75% |
| GHG emissions (million tonnes) | 780 | 845 | 625 | 600 |
| Waste (million tonnes) | 250 | 200 | 175 | 125 |
| Dominant device for online access | mixed | Digital TV | PC/mixed | PC |
| Dominant use of networks | commercial | Commercial | mixed | private |
| Dominant form of e-commerce | B2C | B2B | mixed | B2B |
| POTENTIAL SOCIAL IMPACTS OF THE NEW ECONOMY | | | | |
| Low cost access to online services | ✓✓ | ✓ | ✓✓ | ✗ |
| Access without IT skills | ✓✓ | ✓ | ✗✗ | ✗ |
| Equal access to all online services | ✗ | ✗✗ | ✓ | ✓ |
| Equal opportunities for 'offline population' | ✗✗ | ✗✗ | ✗ | ✓ |
| Strong consumer protection | ✓ | ✗✗ | ✓ | ✓✓ |
| High efficiency of public services | ✓✓ | ✓ | ✗ | ✗✗ |
| Improved political participation | ✓ | ✗ | ✓ | ✓✓ |
| Balanced regional development | ✗✗ | ✗ | ✓ | ✓ |
| POTENTIAL ENVIRONMENTAL IMPACTS OF THE NEW ECONOMY | | | | |
| Resource-efficient production | ✓✓ | ✓ | ✓✓ | ✗✗ |
| Resource-efficient supply chains | ✓✓ | ✓ | ✓✓ | ✗ |
| De-materialisation of products | ✓✓ | ✓ | ✓ | ✗ |
| Sustainable consumption patterns | ✗✗ | ✗ | ✗ | ✓ |
| Green consumerism | ✗ | ✗ | ✓✓ | ✓✓ |
| Less heavy goods transport | ✓ | ✓ | ✓ | ✓✓ |
| Less light goods transport | ✗✗ | ✗ | ✗✗ | ✓ |
| Fewer individual shopping trips | ✓ | ✗ | ✗ | ✗ |

| | |
|---|---|
| ✓✓ | progress towards objective |
| ✓ | some progress towards objective |
| ✗ | no progress towards objective |
| ✗✗ | movement away from objective |

**Table 2-1 Comparing scenarios and impacts (Wilsdon 2001)**

As shown in Table 2-1, the environmental impacts of the e-economy across the scenarios suggested a very mixed picture. None of the scenarios presented an absolute positive or negative result, which indicated the complexity of the relationships.

### 2.3.2 GeSI, 2002

A partner of Digital Europe, the Global e-Sustainability Initiative (GeSI), with the support of the United Nations Environmental Programme (UNEP), aimed to promote ICT as a critical contributor to sustainable development. Research activities included: quantifying the potential contribution of e-business to dematerialisation, resource productivity and eco-efficiency; investigating the relationship between e-business and corporate social responsibility (CSR), assessing the impact of e-business on regional development across Europe.

The major findings of the research was published to a report "Industry as a partner for sustainable development: Information and Communications Technology" in 2002 (GeSI 2002). This report addressed the transformation the telecommunications industry had undergone, and set out the position of ICT in the global economy and how the ICT sector was contributing to the economic development as well as touching on key social issues such as human rights and diversity.

Part 4 dealt with the environmental impacts, provided some statistical data on the contribution both within the ICT industry and in other industries quoted from other studies. The focus was mainly on how the technologies enable more advanced management of environment, for instance through intelligent transport systems, intelligent building management systems, teleworking and teleconferencing.

### 2.3.3 Digital Europe, 2003

Followed the success of Digital Futures project, 3 research institutions in Germany, Italy and the UK joined with 11 international corporations and 2 European regions to explore in greater depth the contribution that e-business and e-work could play in creating a sustainable information society, where came the Digital Europe project, a pan-European research project investigating the impacts of e-business on sustainable development. The Digital Europe project has been undertaken for the European

Commission under the Information Society Technology (IST) programme and made an important contribution to policy development at the European level. This work has coincided with increasing interest in corporate social responsibility and twin commitments from the EU to become the most competitive knowledge economy by 2010 and to integrate sustainable development across all policy areas.

http://www.digital-eu.org site provides reports developed during the project, news about the subject area, interview with leading experts, presentations from the project's final conference, all the final project reports and some case studies.

Digital Europe focused its research on three key themes in e-business: resource efficiency and dematerialisation, corporate responsibility and regional development. Results of the three strands of the project have been published to the final theme reports, correspondingly: virtual dematerialisation and factor X (Kuhndt et al. 2003), social responsibility in the information society (Alakeson et al. 2003), and e-business and sustainable regional development in Europe (DigitalEurope 2003). Other reports looked into issues like environmental and social impacts of mobile computing, digital music, e-banking, and communicating of sustainability guidance for higher education institutions etc.

### 2.3.4 The Impact of ICT on Environment Sustainability, IPTS, 2003

The Future Impact of ICT on Environmental Sustainability project by IPTS (Institute for Prospective Technological Studies, Seville), coordinated by IZT (Institute for Futures Studies and Technology Assessment), Forum for the Future, EMPA (Swiss Federal Laboratories for Materials Testing and Research Sustainable Information Technology Unit), IIIEE (International Institute for Industrial Environmental Economics and Lund University), was to identify corridors of the influence of ICT on the environmental indicators. The study had an explorative character as there were no quantitative studies for the EU yet. Five interim reports were produced from this project: i)'Identification and global description of economic sectors' (Erdmann and Würtenberger 2003); ii)'Script' (Erdmann and Behrendt 2003); iii)'Scenarios' (Goodman and Alakeson 2003); iv)'Refinement and Quantification'(Hilty et al. 2004); v) 'Evaluation and Recommendations'(Arnfalk 2004), and one journal paper (Hilty et al. 2006) with the summary of the results was mentioned earlier in the journal paper review section too.

Prior to this project, the European Commission had launched the eEurope initiative for creating a digital society and building a net economy. The EU strategy for sustainable development was fuelled at the summit in Lisbon (March 2000) by agreeing on a reporting mechanism integrating social and economic indicators. At the summit in Gothenburg (June 2001) the environmental dimension was added, including 6 indicators (greenhouse gas emissions, energy intensity of the economy, volume of transport to gross domestic product, modal split of transport, urban air quality, municipal waste collected, landfilled and incinerated).

The environmental impacts of ICT were divided into three orders:

- First order (direct): The impacts and opportunities created by the physical existence of ICT and the processes involved;
- Second order (indirect): The impacts and opportunities created by the ongoing use and application of ICT;
- Third order: The impacts and opportunities created by the aggregated effects of large numbers of people using ICT over the medium to long term.

Factors (over-arching, to define the scenarios), variables (smaller scale, to define the scenarios too, could be quantified or qualified, change in different ways according to factors), and three scenarios (Technocracy, Government first, Stakeholder democracy) were developed. A simulation model was created after refinement and quantification of the scenarios.

This project studied the ICT impacts on the European economy as a whole from the macro point of view (e.g. average growth rate of ICT waste, percentage of Intelligent Transport Systems of all freight transport, changes of energy consumption reflected on GDP etc).

"The quantity and breadth of information available on the subject of the Internet and electronic commerce is enough to make any researcher or e-business faint. Yet this does not translate to a correspondingly rich analysis of its effects on the environment.

The Internet and other digital technologies which facilitate e-commerce touch upon everything within the realms of the physical and 'virtual': from financial market infrastructure to digital communication. Where, however do environmentally sustainable business and the planet earth fall within the ongoing stream of e-commerce debate? This area is something of a black hole. Our desire for e-commerce-related news stories, statistics, gadgets and comment is implacable but limited in scope."

### 2.3.5 VINNOVA, 2003

A cooperation project was launched under the funding framework of the Swedish Agency for Innovation Systems (VINNOVA) with collaboration of three Swedish universities, Chalmers University of Technology, Linköping University and Lund University to discuss the issue of E-commerce from three different perspectives with an interdisciplinary profile: i) Focus on the flows and distribution systems in supply chains, ii) Focus on the changing roles in marketing channels, and iii) Focus on the packaging and the environmental implications.

Within this project Abukhader (2003) from Lund University has published a Licentiate dissertation with four papers that have been reviewed in the thesis review section of this chapter. As Abukhader commented, the very pilot study made within the proposal of the VINNOVA project indicated there was no established background or even a clear understanding regarding the assessment issue. The theories in the literature seemed diverging and sometimes contradictory, with even confusing use of terms: Internet, E-commerce, Information technology, etc.

Chalmers University of Technology also produced a research paper "e-commerce and Logistical Consequences", in which part 3 analysed how transport and environment changed in connection with e-commerce (Jönson et al. 2000). This sub-project used mainly scenario technique to explore how e-commerce had been changing the logistic, transport and distribution systems based on the Swedish conditions.

In the author's opinion, it was important that the project realised the situation at that early time and put it on the agenda, but only scenarios were raised, no further qualitative/quantitative model or measurement were presented.

## 2.4 Methodologies Review

### 2.4.1 Scenarios

As the major methodology used by the IPTS project - Impact of ICT on Environment Sustainability presented in section 2.3.4, scenarios are qualitative planning and communication tools rather than predictive tools. The project suggested that scenarios were particularly well suited to this area of research since they explicitly acknowledged the complexity and uncertainty that characterised any assessment of the relationship between the development of ICT and environmental sustainability in 2020.

A five step process: identification of variables – identification of factors – building scenario frameworks – validation – developing scenario narratives, was applied in this project. It aimed to provide policymakers and other decision makers with multiple perspectives on key areas of uncertainty and allow them to develop robust strategies that could deal with multiple outcomes. (Goodman and Alakeson 2003)

Also applied in Digital Futures project as the main methodology, scenarios were considered as planning and communication tools to explore complex, uncertain and sometimes disputed futures. Taking the nature of governance and social & political values as two dimensions, the scenarios in this study were generated from a set of conceptual associations, rather than from an empirical model of the real world. The strength of this approach was that it brought an intellectual coherence to the scenarios. However, the two dimensions 'values' and 'governance' used in this project were not simple or easily definable concepts (Wilsdon 2001).

### 2.4.2 Life Cycle Assessment

Life Cycle Assessment (LCA) is a systematic set of procedures for compiling and examining the inputs and outputs of materials and energy and the associated environmental impacts directly attributable to the functioning of a product or service system throughout its life cycle (ISO14040 2006).

LCA, and the economic or environmental Input-Ouput Life Cycle Assessment, has been the most common method applied in this field so far. Just to name a few, Matthews performed a life cycle assessment study for the net effect of environmental implications for book retailing as a case study (Matthews et al. 2001). Caudill et al. also used the traditional life cycle assessment tools and methodologies to exam the impact of e-commerce on the electronic products from the product perspective, using desktop personal computers as a case study. Several assumptions were made in order to quantify and help to understand which lifecycle stages were most important (Caudill et al. 2001). Toffel and Horvath compared reading newspaper on a PDA and the traditional paper format, with the Economic Input-Output LCA for conventional way and traditional LCA for the digital way (Toffel and Horvath 2004).

LCA, being one of the most traditional methods and systematic approach to measure the impacts of certain products/services on the environment, in terms of the energy, raw materials consumption, and final waste in the complete process of production, delivery, transport, consuming and disposal etc, is the most favourable methodology in assessing the environmental impact of e-business/ICT in academics in the author's opinion. Considering the experiences of researchers in applying this method and long established history, LCA is so far the most reliable method too. However, due to the complicated nature of the ICT technology, it is rather difficult to define the boundary of the life cycle and the uncertain/unrecognised factors involved. Most studies are based on quite a few assumptions, restricted in certain narrow field and lack generic application to other sectors.

### 2.4.3 Fuzzy multi-objective optimisation

Luo et al. (Luo et al. 2001) represented and optimised various parameters defining business, technology and environmental issues of the e-supply chain in a fuzzy logic based network to model data uncertainty and information gaps. They proposed a fuzzy multi-objective optimisation method to find the best possible balance between supply chain variables, e-business strategies and environmental performance.

As the case study, an e-supply-chain management of typical desktop personal computers was modelled and optimised. The solutions from the case study yielded some interesting insights regarding performance tradeoffs in e-supply chain networks. "Scenarios of cost and time optimisation suggest to work with partners located close to your facilities and to eliminate intermediate agents between manufacturer and consumers, while energy and emission optimisations seek more environmentally efficient ways to conduct production and logistic".

In the author's opinion, this is one of the most novel approaches applied in this field. However as the work was presented in the form of a short conference paper, details of the application and process are rather limited. This method is definitely worth further exploitation, as the optimisation part seems to have the potential of balancing the environmental impacts and economic benefits resulted from e-business/ICT, which is the ultimate goal of conducting research in this area.

### 2.4.4 Ecological Footprint

Frey et al discussed whether the Ecological Footprint methodology could be applied to electronic products. The Ecological Footprint (EF) methodology, developed by Wackernagel and Rees, has often been suggested as a sustainability indicator for the human impact on earth. EFs, expressed as area, sum up the total productive area of land and water ecosystems required to sustain the resources, wastes, and emissions of a population wherever that land may be located. Thus, EFs can be established on a global or other geographic level.

In this paper, area was used as the single indicator. Frey et al estimated that the EF of a PC is 1790 $m^2$ or 0.18 ha over its lifetime of three years, which was about 9% of the terrestrial EF of a world-average citizen, and it exceeded its own physical size by more than a thousand fold. The results confirmed the use phase as the main culprit, followed by manufacturing and material production. Frey et al proposed that the Ecological Footprint could not compete with other assessment tools, but should rather be seen complementary. (Frey et al. 2000)

### 2.4.5 Geographical Information System

A few conference papers and working papers (e.g. (Cairns 1999), (Punakivi and HolmstroÈ M 2001)) have proposed utilising Geographical Information System (GIS) in managing the environmental impacts of e-commerce or comparing the online system with the traditional system.

Due to their limited academic significance, these papers are not reviewed in details in this thesis. However, GIS is particularly good at route planning in the home delivery in the e-commerce system in the author's opinion, such as how to optimise the route for online deliveries within a geographic area in order to minimise the frequency of delivery and fuel usage etc.

Some industries have started designing and implementing the more systematic GIS based Intelligent Transport System (ITS) for freight transport and to improve the overall supply chain system. The potential of this tool for decoupling the environmental impacts remains to be seen and further researched.

## *2.5 Conference & symposium Review*

### 2.5.1 Symposium at the New York Academy of Science on e-commerce and environment, New York, USA, 24-25 October 2000

The two-day symposium (October $24^{th}$ – $25^{th}$, 2000) sponsored by the New York Academy of Science and the Tellus Institute looked at whether the digital economy would evolve into a powerful ally of the earth's vulnerable environment or jeopardise it even further (New York Academy of Science 2000). The keynote speakers included Braden Allenby, vice president of Environment, Health and Safety from AT&T, James Wilsdon, senior policy adviser from Forum for the Future, UK, and David Festa, assistant to the secretary and director of the Office of Policy and Strategy Planning, US department of Commerce. Other prominent speakers included senior economist or executive directors from US Environmental Protection Agency and other government bodies as well as top universities. The high profile speakers brought this issue up to an unprecedented level and importance, not only to academia, but also to the industry and government in United States.

Topics addressed at this symposium included:

- E-commerce and product/process design: opportunities for energy and material efficiency in the design of products and e-commerce systems. Would e-commerce make business more energy efficient or put a strain on US power plants? Did the digitalisation of products-books, music, and photos - minimise or create waste?

- E-commerce and logistics: the possibility of maximising transportation and energy efficiency and minimising waste throughout the supply chain. How could we improve the movement of goods in the digital economy, from industrial freight to books ordered online, so that it was energy-efficient, minimises packaging, and averts congestion?

- E-commerce and land-use: the potential for minimising sprawl and maximising opportunities for smart growth. How would our communities be affected by business moving online? Would downtowns be hurt? Would small communities be helped? Would new delivery systems reduce or increase traffic?

### 2.5.2 The 2001 IEEE International Symposium on Electronics and the Environment, Denver USA, 7-9 May 2001

The IEEE Computer Society, Technical Committee on Electronics and the Environment sponsored the 2001 IEEE International Symposium on Electronics and the Environment in Denver, Colorado (IEEE 2001). The theme "Life-Cycle Environmental Stewardship for Electronic Products" reaffirmed the electronics industry's leadership in environmental design and management, including life-cycle management, disassembly, recycling, takeback, energy use minimisation, elimination of hazardous substances, supply chain management, and other activities, also covered some new domains such as e-commerce, new materials and components, and advances in manufacturing.

There were fifteen sessions/sub-themes in the proceeding, just to name a few, session 1 entitled "DFE Methodology & Tools I" including paper "Eco-Labelling and the Information Technology (IT) Industry" (Hermann et al. 2001); session 9 entitled "analysis of the Impacts of E-everything", including paper "The Net Effect: Environmental Implications of E-commerce and Logistics" (Matthews et al. 2001);

session 14 entitled "Product Reuse & Recycling" including paper "Environmental Impact of a Telecommunication Service" (Taiariol et al. 2001). The proceeding can be accessed online at

http://ieeexplore.ieee.org/xpl/tocresult.jsp?isnumber=19985&isYear=2001# .

### 2.5.3 Joint OECD/ECMT Seminar on the Impact of E-commerce on Transport, Paris, France, 5-6 June 2001

On 5-6 June 2001, the OECD and ECMT held a joint International Seminar on "The Impacts of E-commerce on Transport" (OECD/ECMT 2001). The aim of this Seminar was to analyse the relationship between e-commerce and transport and to identify the likely consequences of e-commerce on transport specifically.

Conclusions were drawn as "It is important to be cautious in drawing conclusions from a superficial review of the links between e-commerce and transport. There is a lack of concrete information available on e-commerce. This is a concern that is reflected in the numerous research studies currently being conducted. Governments need to realise that e-commerce may result in increased freight transport demand without reduction of passenger transport demand, if necessary measures are not taken."

"More information on market segmentation is required in order to conduct detailed field analyses and gather reliable statistical data on urban deliveries and the impact that e-commerce has had on them. As well as this, consumer behaviour, expectations, life-style and reactions to the early stages of e-commerce are other relevant areas for in-depth investigation. Innovative, forward-looking approaches are warranted to distinguish short-term marginal impacts from long term impacts, which are unpredictable, but may be major."

"Given the great uncertainty and seriousness of long-term impacts of e-commerce on transport, it is important to monitor systematically the developments of positive and negative impacts of e-commerce and whether governments' objectives are being met."

## 2.6 Organisations & Websites

Websites are in general excluded from PhD literature reviews, as an unconventional form of academic publication as anybody can set up a world-wide-web to express any personal opinions to the public without acknowledgement or reference. However due to the nature of this study having ICT/e-business unavoidably linked with websites and the fact that figures/facts/analyses quoted by many journal papers actually come originally from these websites, a review is therefore included. The author believes that any form of information is important at the start of a new research topic.

### 2.6.1 http://www.svtc.org/


**Figure 2-2 SVTC website**

Silicon Valley Toxics Coalition (SVTC) is a diverse grassroots coalition that engages in research, advocacy, and organising around the environmental and human health problems caused by the rapid growth of the high-tech electronics industry (SVTC 2006).

SVTC was formed in 1982 in response to the discovery of substantial groundwater contamination throughout Silicon Valley that was caused by toxic chemicals that leaked out of underground storage tanks from high-tech companies (at the time referred to as the "clean industry!")

Since then, SVTC has expanded its programs to address a wide variety of sustainability issues associated with the high-tech electronics revolution. The program efforts spring from a philosophical commitment and our long track record supporting community involvement and environmental and economic justice.

The website provides some interesting resources, for example, 34 site-level interactive (point-and-click) maps of Santa Clara County that illustrate the 150+ ground water contamination and 29 Superfund sites throughout Silicon Valley, and some interesting facts/figures which have been frequently quoted by many papers in regard as the energy/resource used in ICT hardware production, waste generated and contamination caused, such as "800 kWh of electrical energy is consumed in the manufacturing of a single 200 mm semiconductor wafer. This is enough energy to supply a typical household for 2 months".

### 2.6.2 http://www.digital-eu.org/

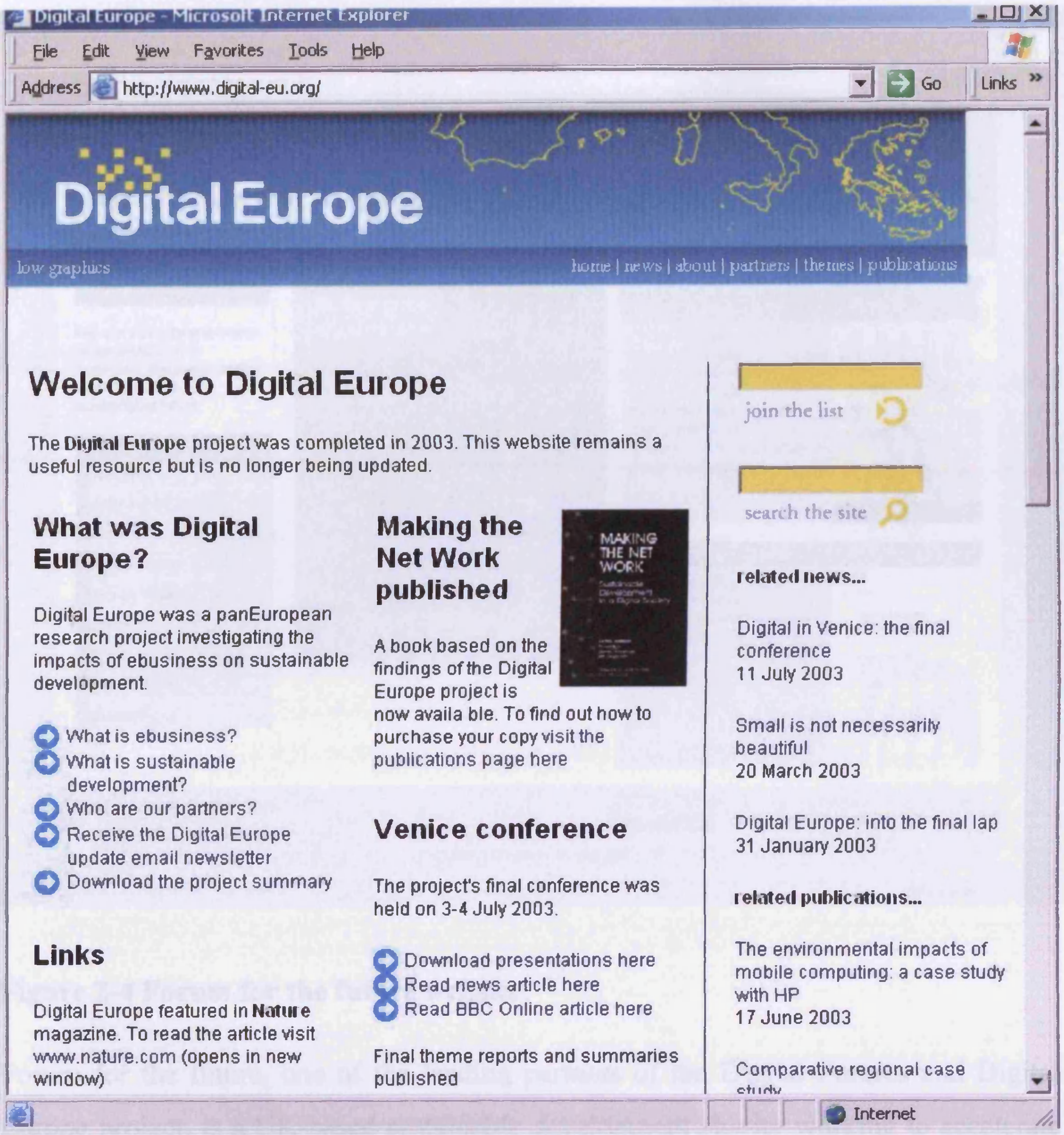



**Figure 2-3 Digital Europe website**

This website provides reports developed in the Digital Europe project which has been reviewed earlier in this chapter, news about the subject area, interviews with leading experts, presentations from the project's final conference and all the final project reports, free to download (Digital-EU 2006).

### 2.6.3 http://www.forumforthefuture.org.uk/

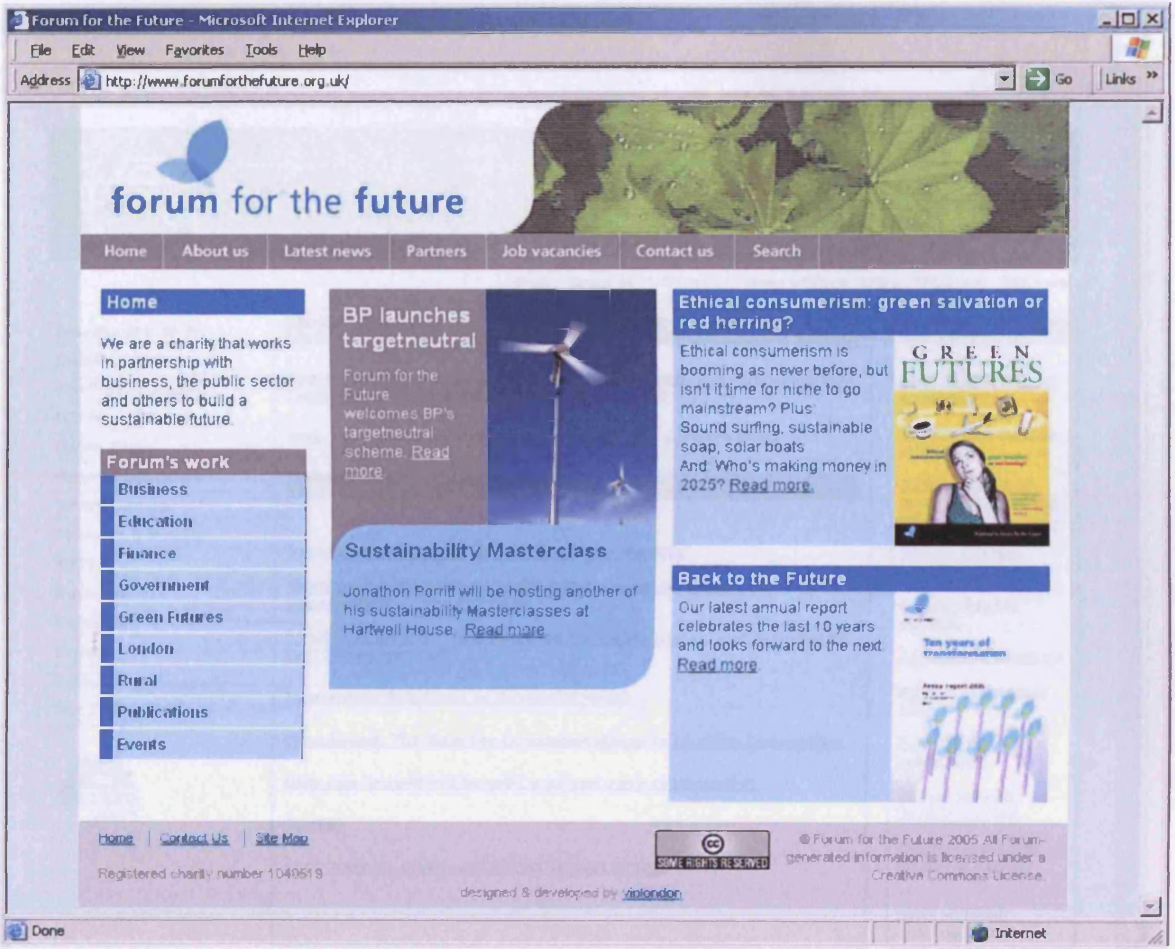



**Figure 2-4 Forum for the future website**

Forum for the future, one of the leading partners of the Digital Futures and Digital Europe project, is a UK-based sustainable development charity working to accelerate the transition to a sustainable way of life. It provides advice and develops partnership work on issues as diverse as climate change, procurement strategies, environmental accounting and the digital divide (Forum for the future 2006).

The website provides models of Sustainable Development, namely are the triple bottom line, the five capitals model and the twelve features of a sustainable society.

### 2.6.4 http://www.sustainit.org/

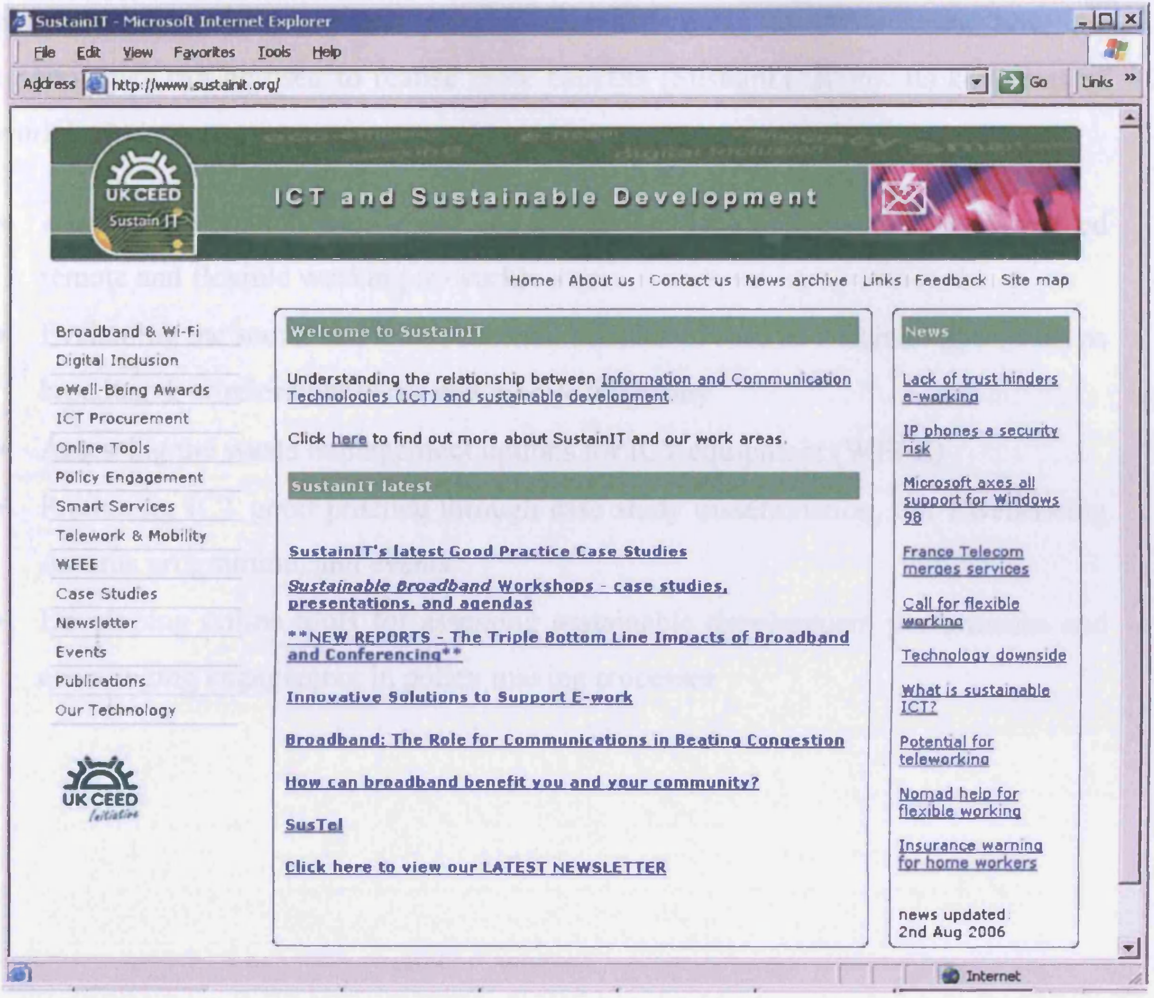



**Figure 2-5 SustainIT website**

SustainIT, an initiative of leading charity UK CEED, focuses on the relationship between ICT and sustainable development and seeks to demonstrate how the technologies can be used to realise these benefits (SustainIT 2006). Its key areas of work include:

- Assessing the economic, social and environmental implications of ICT-enabled remote and flexible working (e-work)
- Evaluating the social and environmental benefits of core ICT technologies, such as broadband, wireless networks and mobile telephony
- Assessing the waste management options for ICT equipment (WEEE)
- Promoting ICT good practice through case study dissemination, our Ewell-Being Awards programme, and events
- Developing online tools for assessing sustainable development performance and encouraging engagement in policy making processes

### 2.6.5 http://www.imrg.org/


Figure 2-6 IMRG website

IMRG is the leading industry body for global e-Retailing. It tries to talk to academics and other interested parties around the world to assess what research has already been completed and what still needs to be done in order to understand the overall effect of Internet shopping on carbon emissions. Its Industry Statistics section presented some whitepapers such as Effects of e-commerce Green House Gas emissions, the Impact of ICT on Travel and Freight (IMRG 2006).

### 2.6.6 http://www.greenhome.com/

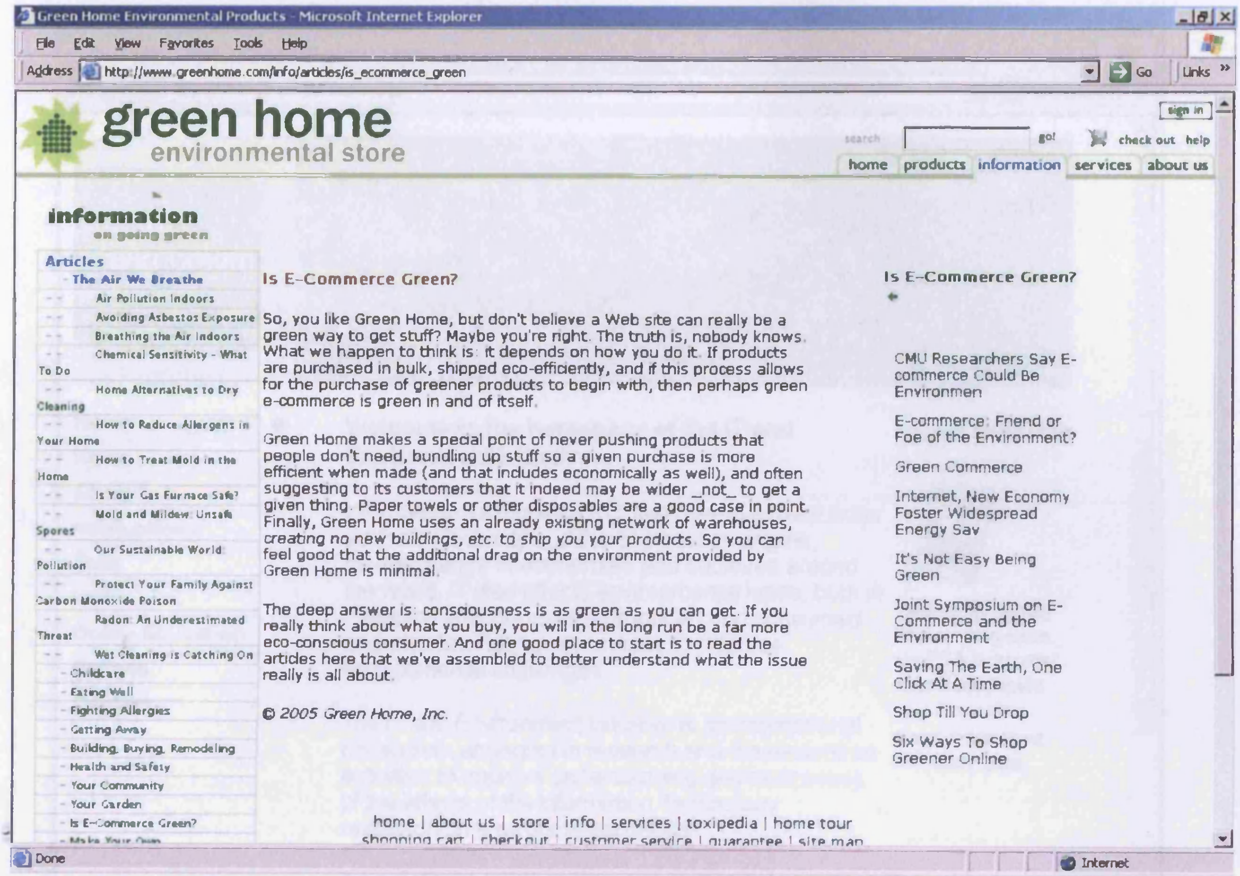



**Figure 2-7 Green-home website**

This website offers environmental products as well as information sources, including some articles frequently quoted by many papers, such as "Is E-commerce Green?", "E-commerce: Friend or Foe of the Environment?", "Saving the Earth, One Click at A Time", " Six Ways to Shop Greener Online" etc (Green-Home 2006).

### 2.6.7 http://www.it-environment.org/

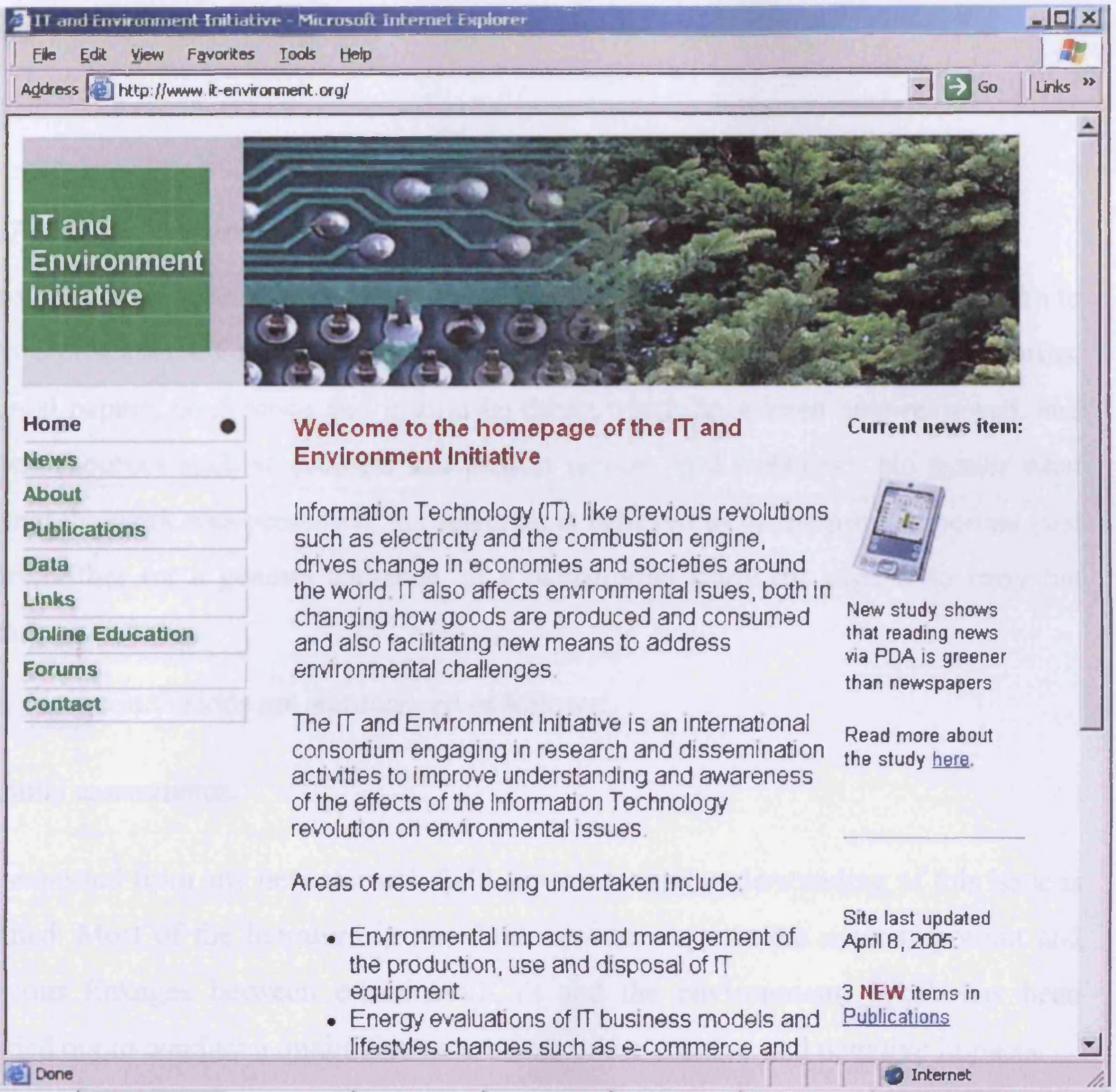



**Figure 2-8 it-environment website**

The IT and Environment Initiative is an international consortium engaging in research and dissemination activities to improve understanding and awareness of the effects of the Information Technology revolution on environmental issues, collaborated by 3 Japanese universities and Carnegie Mellon University in US (IT and Environment Initiative 2006).

Areas of research being undertaken include: environmental impacts and management of the production, use and disposal of IT equipment; energy evaluations of IT business models and lifestyles changes such as e-commerce and telecommuting; macro-economic analysis of the relationship between IT investments and energy use at the national level; analysis of effects of IT-driven shifts in lifestyles and consumption pattern on embodied energy in consumption.

## 2.7 Conclusions

This chapter has attempted to provide a review of the published literature in relation to how e-business/ICT affects the environment. The work reviewed is in various forms, journal papers, conference and symposia, thesis which have been peer-reviewed, and other resources such as projects and project reports, and websites. No matter what format the work was presented, the selection is believed to be the most important past work, either for a general audience, or a background study for experts to carry out future research.

The main conclusions are summarised as follows:

i) Initial assessments.

As expected from any new research field, awareness and understanding of this issue is limited. Most of the literature in this field aims to map out the most important and obvious linkages between e-business/ICTs and the environment. Work has been carried out to conduct a qualitative assessment of the positive and negative impacts.

Past works have looked at the problem from limited specific segments, e.g. teleconferencing technologies and transport, or e-commerce and packaging waste. Few attempts have been made to look at the whole picture. Researchers agree that the problem is not straightforward and is worth further investigation. The various contributions to date have helped shape a promising and interesting research path.

Studies so far present a mixed picture and conclusions. But the more important thing is, academia and government bodies, have begun paying attention to this problem domain and started researching.

ii) Methodologies.

Methodologies applied in or proposed by previous studies have been reviewed and categorised in a separate section due to their importance for future research. A few works have attempted to either extend an existing methodology or develop a new method to solve the problem. Traditional methods and a systematic approach for impact studies have been found to be insufficient for this research topic due to its complicated nature.

iii) Indicators and models.

It is acknowledged that there are already studies, which develop quantitative indicators. However, the current research community at the stage is trying to develop a perfect set of indicators for a specific field. Having realised that indicators are just a tool to analyse the overall relationship, so instead of joining the main theme, this PhD study attempts to take one step ahead to look at what can be done generically to the indicators for exploring/extracting further information and knowledge in the future, and how this could be achieved.

A few models have been built to attempt to simulate certain specific parts of the relationship. However, these models have been developed based on a number of assumptions and have not yet been fully validated for major applications. In the author's opinion these are a good start but still need further consideration. A more predictive and empirical model, which can be applied within a sector of society, should be more beneficial in the long term. Such an approach should help simulate potential impacts resulting from changes of indicator, so that positive effects can be promoted and negative ones alleviated. This PhD is directed towards that goal.

iv) Meso-level approach.

The currently dominant approach is either a micro-level case study approach or a macro-level statistical approach, both only presenting in general terms the comparison results or outcomes of the impacts. The macro statistics approach provides a general picture which says whether ICT/e-business is "good" or "bad" for the economy, society and the environment. However the approach has been unable to answer what action is needed to promote the good and avoid the bad, in relation to each component (individual person, company, industry and sector) in the society. On the other hand, micro level case studies offer limited contribution due to their restriction to certain specific products, companies, geographic areas etc.

In the author's opinion, industry or sector based meso-level studies have the potential to not only present the good/bad statistics within an industry/sector, but also offer a solution of deciding what to change, to achieve certain goal/objectives in the industry/sector. Meso level studies do not exist in this research field at the moment. In the author's opinion they may be able to fill the gap and contribute to understanding and control of the problem significantly. This study will also attempt to explore this area.

# Chapter 3

# Data Collection and Data Analysis

## *3.1 Introduction*

In order to analyse the relationship between e-business/ICT and the environment, the approach adopted in this study was to simulate how e-business indicators interact with environmental indicators quantitatively. To achieve this aim, the first step necessary was to collect relevant data that represent indicator values from various sources. This chapter will provide details of that data collection progress and preparation, at the initial stage of the study, as well as the qualitative analysis and assessment of the results.

Due to lack of data and information in the current area in government bodies/councils/ research institutes, two questionnaire surveys were conducted in two industry sectors.

The following sections will explain in detail

- The choice of the two industry sectors selected
- The design of the questionnaires in terms of content and format
- The storage of the data, in two databases respectively
- The pre-processing of the data, including data normalisation, choosing indicators
- The qualitative analysis and the statistical analysis of the results
- The conclusions drawn from the two surveys respectively

## 3.2 Choice of Industry

It was critical for this study to choose the right industry to survey. The industry needed to be both environment conscious and active in utilising ICT technologies or an e-business platform, so that it would be feasible to demonstrate the possible linkage between these two, no matter whether it was positive or negative. After a literature study, comparisons, assessment and talking to some experts and industrial links, it was decided to conduct the survey in the top 500 retail companies within UK.

The editor of the UNEP (United Nations Environmental Program) magazine *Industry and Environment* (Editorial 2003) stated that, in recent years a few companies in the retail sector not only started to green their own operations but also became important players in global efforts to make consumption and production patterns more sustainable. They were taking action: developing logistical strategies for transport, making life-cycle assessments of packaging, adopting environmental management systems (EMS), marketing green products, ensuring that sustainability was taken into account in supply chain management, drawing up codes of conduct for suppliers, carrying out consumer and employee awareness-raising campaigns, demanding innovation in building design and energy systems.

As Durieu stated, the challenge of sustainable retailing was twofold: (i) to ensure that internal operations were sustainable, and (ii) to influence suppliers to produce sustainably and customers to consume sustainably. By acting as a go-between for consumers and producers, communicating the demands of customers upstream to their suppliers and delivering new products and services downstream to their customers, retailers played a critical role in shaping production processes and consumption patterns (Durieu 2003).

Therefore retail sector satisfied being an influential role in promoting environmental sustainability due to its strong involvement with large number of other industries, what about the ICT diffusion in this sector?

For the EU4 (D, I, UK, F) more than 80% of the enterprises in retail sector had access to the Internet, above all on web protocol (72.8%). Such percentage decreased considering the enterprises that use the web as an instrument of communication or e-commerce (57.5% of the EU4), the highest peak was represented from UK, where 3 out of 4 of the retail's enterprises had their own website. The impact of online sales with respect to total company sales was less than 5% in most cases in EU, but in UK approximately one-quarter of the companies that sell online generated over 50% of their turnover from these sales (EC 2002). The retail industry was doing very well with the ICT diffusion, UK in particular. Moreover, it was found that it was the largest retailers who were at the forefront of the e-commerce revolution in the UK (Ellis-Chadwick et al. 2002).

| | EU4 | D | EL | F | I | NL | A | P | UK |
|---|---|---|---|---|---|---|---|---|---|
| Access to the Internet | 82.0 | 83.4 | 77.4 | 65.8 | 87.5 | 71.8 | 91.5 | 32.6 | 88.1 |
| World wide web | 72.8 | 75.1 | 39.4 | 56.2 | 68.4 | 68.5 | 86.9 | 21.4 | 83.9 |
| Website on the Internet | 57.5 | 58.9 | 34.1 | 27.8 | 58.1 | 71.3 | 66.2 | 12.6 | 75.6 |
| Intranet | 41.7 | 30.8 | 44.6 | 56.4 | 34.4 | 40.7 | 48.8 | 15.1 | 48.4 |
| Extranet | 12.5 | 9.0 | 3.5 | 21.2 | 5.7 | 15.2 | 14.9 | 9.4 | 14.6 |
| Local Area Network | 54.3 | 50.2 | 65.3 | 47.4 | 38.9 | 38.7 | 54.0 | 15.2 | 72.6 |
| Wide Area Network | 28.5 | 21.4 | 23.8 | 21.5 | 10.2 | 25.1 | 19.2 | 10.8 | 52.1 |
| Electronic Data Interchange | 20.3 | 16.7 | 14.6 | 40.5 | 3.0 | 22.6 | 22.0 | 11.5 | 20.8 |
| Computational base: all enterprises | | | | | | | | | |

**Table 3-1 Usage of network applications (% of enterprises, 2002) (EC 2002)**

Thus the retail sector has demonstrated to be a strong player in both environment protection and ICT/e-business taking up, therefore it was chosen to be the industry to survey.

However the UK retail market was dominated by a few large chain companies such as Tesco, Sainsbury and Asda, rest of the industry was consisted of mainly small and medium sized companies (SMEs). As a relatively small country (in terms of population and area), companies at the middle to bottom of the top 500 ranking were expected to be small sized shops that many of them could hardly care about the environment or reach the digital platform in reality thus could provide little value to this research. But in order to maximise the return rate, questionnaires were sent to all without exclusion. One industrial expert suggested taking the survey to a wider audience to include other countries in the EU, but considering the results might be different among countries and also due to the higher cost of survey, the idea couldn't be taken forward at the time.

Retail week, UK's top source of retail news and analysis, produced a book entitled Retail Week's Top 500 (Clark et al. 2001), which provided profiles of the top 500 retail companies (by operation) in the UK. This study used this list as a starting point for the survey.

## *3.3 Questionnaire design*

Prior to designing the questionnaire, choosing the right industry to conduct the survey was taken importantly. Questionnaire was not considered as an effective means to collect data, as there were no incentives involved and also due to lack of awareness and monitoring of certain data by industries. However, it was the only resort available at the time. It was hoped to achieve a reasonable response rate so that basic simulations and models could be built to demonstrate the idea.

### 3.3.1 Contents of the questionnaire

The ultimate goal of this survey was to find out whether or not and how e-business/ICT interacts with the environmental performance in a company within the retail industry. Data was to be collected on all e-business/ICT related issues, environmental acts along with basic company profile information, so they could be fed into the Neural Network to be analysed thus a conclusion could be made to tell whether a relationship existed and how changes on one side could affect the other.

This survey was designed to find out:

- How companies' attitudes toward the environment affected their business performance. (Whether proactive environment protection helped companies improve their business performance)

- How e-commerce contributed to a company's environmental performance (Whether the more e-commerce adoption improved or deteriorated environment, including energy use, transportation needs etc.)

- How pro-environment requirements from government affected a company's business? (Add on pressure? / Opportunities? / Threats?)

Accordingly the questionnaire was designed in seven sections:

i) Contact information
ii) Company profile (including employee number, age of the company, number of outlets etc)
iii) The company's environmental policies, legislation and systems (a company's general attitude and policy towards environment protection and systems developed in place)
iv) Resources and outcome of environment protection actions (data collection on energy consumption, material usage and disposal, and financial outcome etc)
v) Human Resource Management and the environment (environment awareness training issues)
vi) E-business taking up information
vii) Interaction of e-business and the environment (direct picturing of the relationship before data analysis from the retail industry's point of view)

In order to choose the right questions to ask, literature review in questionnaire survey techniques and designs was conducted (e.g. (Fink 1995)), also what companies currently report actively were looked into.

### 3.3.2 Environmental performance reporting

It was recognised that in October 2000 the Prime Minister challenged the top 350 UK companies to produce environmental reports by the end of 2001, which was likely to become a compulsory requirement for all companies eventually in the near future. So more and more companies started to produce or prepare environmental reports, however with various purposes: to enhance the organisation's reputation as a responsible employer and supplier, differentiate the company, gain access to capitals, attract more customers, aid communication with stakeholders or help identify opportunities and manage risks.

Nevertheless bearing different purposes, the companies reported their environmental performances in all different ways from various aspects. Some companies produced standalone environmental reports, some produced as part of a wider Corporate Social Responsibility (CSR) reports, and some embedded in the annual report with a mesh mash of everything. In terms of the content related to this issue, they looked into different things and expressed in all kinds of ways too: some gave qualitative comparison results of two recent years; some presented absolute quantitative figures directly, however with different units (money wise or energy units directly), some reported already normalised data...

During the process of the investigation, it was realised a standardised report scheme with pre-defined methodology and indicators and practical guidelines were urgently needed. After years' practice and uniformed guidance on producing annual financial reports, companies have unified the format of reporting such as cash flow, balance sheet. But environment reporting is still at infant stage, guidance at the policy and regulation level exists but needs to be widely promoted to the end-users.

The Global Reporting Initiative (www.globalreporting.org) published general guidelines for company sustainability reporting on social, environmental and economic issues. The UK Department for Environment, Food and Rural Affairs (DEFRA) has also published a *General Guidelines on Environmental Reporting* (DEFRA 2001), which in the author's opinion, is an excellent practical guidebook for industries to start reporting Greenhouse gas emission (DEFRA 2000a), waste emissions (DEFRA 2000c) and water use (DEFRA 2000b). The report explained why, when and how should a company produce an environment report as well as what to report. Its example environmental performance indicators and flowchart are straightforward to follow and use. However awareness of the guidebook needs to be raised and probably corresponding trainings should be provided too. This guideline not only helped this study design the questionnaire but also provided the conversion and standardisation methods for the data analysis at later stage.

Anyway after summarising the common points, questions were drawn up within the seven areas mentioned above. Principle environmental problems include waste, packaging, paper waste, energy consumption, land contamination, EMS, environment auditing procedure, environment related training schemes and management issues, joining ISO 14001 (ISO 2006) and EU eco-label Scheme (Eco-Label 2006), and the main reasons of pro-environment actions. E-business related questions include: online sales & purchases percentage, history of using electronic network, electronic media applied, website, teleconferencing and telecommuting. The linkages section asked about the financial input and results of pro-environment. And the last section intended to see how general audience view the relationship between e-business/ICT and the environment.

### 3.3.3 Cover letter and questionnaire

The original cover letter and questionnaire are presented in Appendix A.

## *3.4 Survey 1 Process*

Questionnaire 1 *Survey of "Relation between E-commerce and the Environment"* was sent out on October 2003 to the top 500 retail companies in UK by post. A decision regarding the addressee of the letter had to be taken. It was found that for some companies their health and safety managers dealt with environmental issues. In those cases the questionnaires were sent to the health and safety manager. Otherwise it was sent to IT managers if applicable, or CEOs who were expected to redirect the questionnaire to whoever was appropriate in the company.

Follow-up emails were sent and calls were made after two weeks' posting, as the returns were rather poor. However the replies were rather unenthusiastic. The main reasons were that most companies didn't measure these problems so they didn't have or they were not aware of any data available. Another realistic problem was it would be rather difficult for an environmental manager to answer IT related questions. On the other hand it wouldn't be easy for IT managers to be aware of any environment related data.

## *3.5 Results and Analysis of Survey 1*

The survey officially finished on February 2004. By then the response rate was 7.25%, while completed questionnaire rate that provided actually complete data was only 2.94%. Other approaches were tried to complement the result, such as extracting data from some companies' environmental report, CSR report and annual report.

### 3.5.1 Data Normalisation

Electricity and gas consumption were standardised in same units and converted to $CO_2$ equivalent according to DEFRA guidelines (DEFRA 2000a) (Table 3-2).

## Table 1: Using the conversion tables

**1.** Put in here the amount of fuel used

**2.** Multiply the amount of fuel used by the conversion factor to get the total $CO_2$ emissions produced

**3.** If there is more than one unit for measuring fuel, then specify here

**4.** Use the relevant factor for converting, in this example, therms of gas used into $CO_2$

| Common Fuel types | Amount used per year | Units | x | Factor | Total kg $CO_2$ |
|---|---|---|---|---|---|
| Grid electricity | 50000 | kWh | x | 0.43 | *21500* |
| Natural gas | | kWh | x | 0.19 | *55000* |
| | 10000 | **therms** | x | **5.50** | |
| Gas/diesel oil | | tonnes | x | 3142 | |
| | | kWh | x | 0.25 | |
| | | litres | x | 2.68 | |
| Petrol | | tonnes | x | 3135 | *138600* |
| | | kWh | x | 0.24 | |
| | 60000 | litres | x | 2.31 | |
| ***Aggregate total emissions from energy use*** | | | | | *215100* |

**5.** Add the totals for $CO_2$ to get the aggregate total for your energy use.

**Table 3-2 Example DEFRA $CO_2$ conversion table (DEFRA 2000a)**

Most data collected were for 2002-2003 as the most recent financial year when this PhD research started. Energy consumption figures were asked in either KWh unit or the monetary form for convenience of answering but were standardised to unit MWh later, data provided in sterling pounds was converted using British Gas (2003 scale) price as proxy value. Electricity consumed was assumed to be grid electricity rather than from renewable energy.

### 3.5.2 Data Storage

A Microsoft Access database was built to store the data (Figure 3-1).

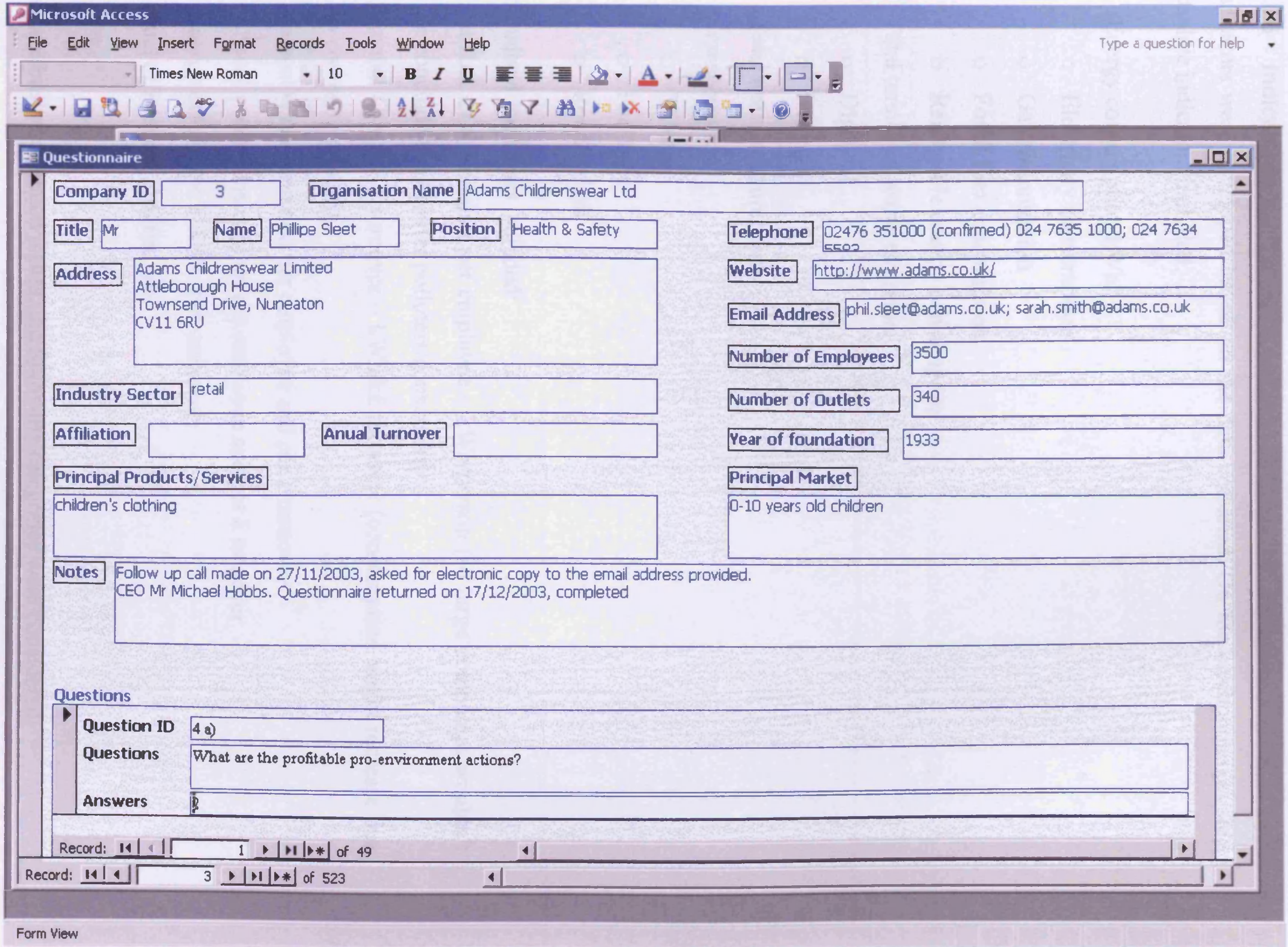



**Figure 3-1 Database for survey 1**

### 3.5.3 Indicators

Indicators were developed based on the data collected. On the environmental side, absolute indicators included:

- Energy consumption kWh
    - Electricity consumption
    - Gas consumption
    - Fossil fuel consumption
    - Renewable energy consumption
- Total tonnes of waste emission
    - Disposal
    - Recycle
- Total water consumption
- Transportation
    - Mileage
    - Fuel type
    - Road/air/rail

Normalised indicators included:

- Annual $CO_2$ emission per employee – kWh/person (so large companies wouldn't be concluded as heavier polluters/consumers)
- Annual $CO_2$ per £ turnover – kWh/k£ turnover (consumption might increase due to company growth)
- Annual waste emission per employee and per £ turnover
- Annual water consumption per employee and per £ turnover
- Energy efficiency: kWh/£ '000 turnover
- Fuel efficiency: km/litre fuel

On the e-business related issues, the following categories were considered:

- Teleworking related energy and cost factors (energy save, cost save etc are presented as minus, energy consumption and GHG emission etc are presented as plus)

| | Number (person) | Office Space (square feet) | Net Energy Save (MWh) | Cost Save (£) | CO2 reduced (kg) |
|---|---|---|---|---|---|
| Tele-worker | a | -175 * a | - 3 * a | -110 * a | - 3000 * a * 0.43 |
| Home-worker | b | -300 * b | - 4 * b | -147 * b | - 4000 * b * 0.43 |

**Table 3-3 Teleworking conversion (Keskinen et al. 2001)**

- Business travel $CO_2$ emission and cost saving through Videoconferencing (per trip)

| | Number of persons | Distance (km) | Cost Total(£) | Meeting time (hours) | CO2 (kg) | Cost saving (£) |
|---|---|---|---|---|---|---|
| Average Petrol car | P1 | D1 | C1 | H1 | + 0.20 *P1* D1 – 1.5265 * H1 | 50 * H1 – C1 |
| Average diesel car | P2 | D2 | C2 | H2 | + 0.12 *P2* D2 – 1.5265 * H2 | 50 * H2 –C2 |
| Rail | P3 | D3 | C3 | H3 | + 0.06 *P3* D3 – 1.5265 * H3 | 50 * H3 – C3 |
| Air | P4 | D4 | C4 | H4 | + 0.18 *P4* D4 – 1.5265 * H4 | 50 * H4 – C4 |
| European Air | P5 | D5 | C5 | H5 | + 0.18 *P5* D5 – 1.5265 * H5 | 80 * H5 – C5 |
| International Air | P6 | D6 | C6 | H6 | + 0.11 *P6* D6 – 1.5265 * H6 | 95 * H5 – C6 |

**Table 3-4 business travel conversion**

The electricity consumption of tele-conferencing was calculated as 3.55kW per hour according to

| | | | |
|---|---|---|---|
| T135 | Electricity Consumption Factor for Virtual Meetings per ph | Direct electricity consumption caused by virtual meetings, including client-, server- and network-type devices, in kWh per telework hour = kW. | 3.55 kW |

**Table 3-5 Electricity consumption factor for Virtual Meetings (Hilty et al. 2004)**

And per kilo watt electricity (grid electricity assumed) would produce 0.43 kg $CO_2$ (DEFRA 2000a), thus a teleconferencing meeting lasted H hours was supposed to produce

*3.55kw/hour * H(hour) * 0.43 kg CO2/kWh = 1.5265 * H (kg) $CO_2$ emission.*

Standard charge of video conferencing was calculated as (University-of-Abertay-Dundee 2006) :
Commercial hire using ISDN dial-in/out: £50/hour
ISDN2 UK -- £10/hour; Europe -- £80/hour; International -- £95/hour

- Number of Computers in the Company (N)

| | Energy used (MWh) | $CO_2$ (tonnes) | Waste (tonnes) |
|---|---|---|---|
| Production | +1.00 * N | +0.19 * N | +0.036 * N |
| Use phase | +2.78 * N | +0.45 * N | +0.108 * N |

Cost: £800 * N

**Table 3-6 Computer related energy use, $CO_2$ and waste**

Source: The use phase of a personal computer in the EU has been calculated to produce 0.45 tonnes of greenhouse gases, 108kg of waste and 10 GJ of energy per unit.

This was much higher than the amounts created during the production of the PC in the first place: 0.19 tonnes of greenhouse gasses, 36kg of waste and 3.6 GJ of energy per unit. (Atlantic-Consulting 1998)

### 3.5.4 Conclusions

Among the very limited data collected, it was found that:

- Most companies believed protecting the environment was important, but not at the cost of sacrificing profit.
- Some felt that e-business and Internet technologies contributed negatively to the environment, because of staff printing emails.
- Some percentage of the industry was using EDI rather than the Internet.
- Around half of the companies had EMS in place. For companies who hadn't done so, the main reason was there was no in-house expert to address the related issues.
- Most large size companies did not actively participate in the survey.

- Some oil and gas companies took environmental issues very importantly, and believed e-business and Internet technologies could help protect the environment. Most of them had already implemented telecommuting, teleconferencing, renewable resources etc

In terms of the data itself,

- Availability of data was limited. The industry was very reluctant to cooperate. Collecting data had been a barrier for the study.

- There was some confusion in the limited data provided. For example, some provided electricity consumption in office (headquarters) while others included outlets. Also some provided data for a specific branch while others offered world wide offices as a total.

The quantity of the data collected was rather insufficient and the quality was also too poor to build any model or run any simulation. Unfortunately, this limited amount of data provided little value for further research. It was therefore decided to seek further data. The approach adopted is described below.

## 3.6 Questionnaire Design - Survey 2

Limited data from the first survey caused significant difficulty for the study to proceed. It was therefore decided that further data would need to be collected, and that a second survey would be required. It was decided that a second survey would be sent to members of the LRN (GRC 2006) project within the author's research centre. It was expected that these members would be more cooperative.

The questionnaire was revised to avoid questions that were found to be difficult to measure from the first survey. How the questions are related to and contribute to measuring the environmental indicators and e-business indicators are presented in the following table.

| Qs | $CO_2$ | Waste | Land | e-business | Environmental activeness |
|---|---|---|---|---|---|
| How many computers in the company? | √ | √ | | √ | |
| Electricity consumption (KWh) | √ | | | √ | |
| Gas consumption (KWh) | √ | | | | |
| Warehouse space ($m^2$) | √ | | √ | √ | |
| Office space ($m^2$) | √ | | √ | √ | |
| Waste disposed (tonnes) | | √ | | √ | |
| Waste recycled (tonnes) | | √ | | √ | √ |
| How many tele-workers? | √ | √ | √ | √ | √ |
| Tele-conferencing? How much persons hours per year? | √ | | | √ | √ |
| Business travel: distance traveled by air (Km) | √ | | | √ | |
| Business travel: distance traveled by train (Km) | √ | | | √ | |
| Business travel: distance traveled by road (Km) | √ | | | √ | |
| Online sales delivery | √ | | | √ | |
| Paper consumption up or down after EC adoption | | | | √ | |
| Implemented EMS/ISO14001? | | | | | √ |
| Employee number | √ | √ | √ | √ | √ |
| Annual turnover/... | √ | √ | √ | √ | √ |
| Environment report / CSR report? | | | | | √ |
| Age of company (year founded) | | | | √ | √ |
| Percentage of online sales | | | | √ | |
| Percentage of online procurement | | | | √ | |
| EDI? | | | | √ | |
| Web presence? | | | | √ | |

Table 3-7 Questionnaire factors

## 3.7 Questionnaire 2

The original cover letter and questionnaire are presented in Appendix B. An online version of the questionnaire (Figure 3-2) was designed in web page format, as a complementary method for collecting data. This proved to be much more successful than the traditional mail or emailing attachments.

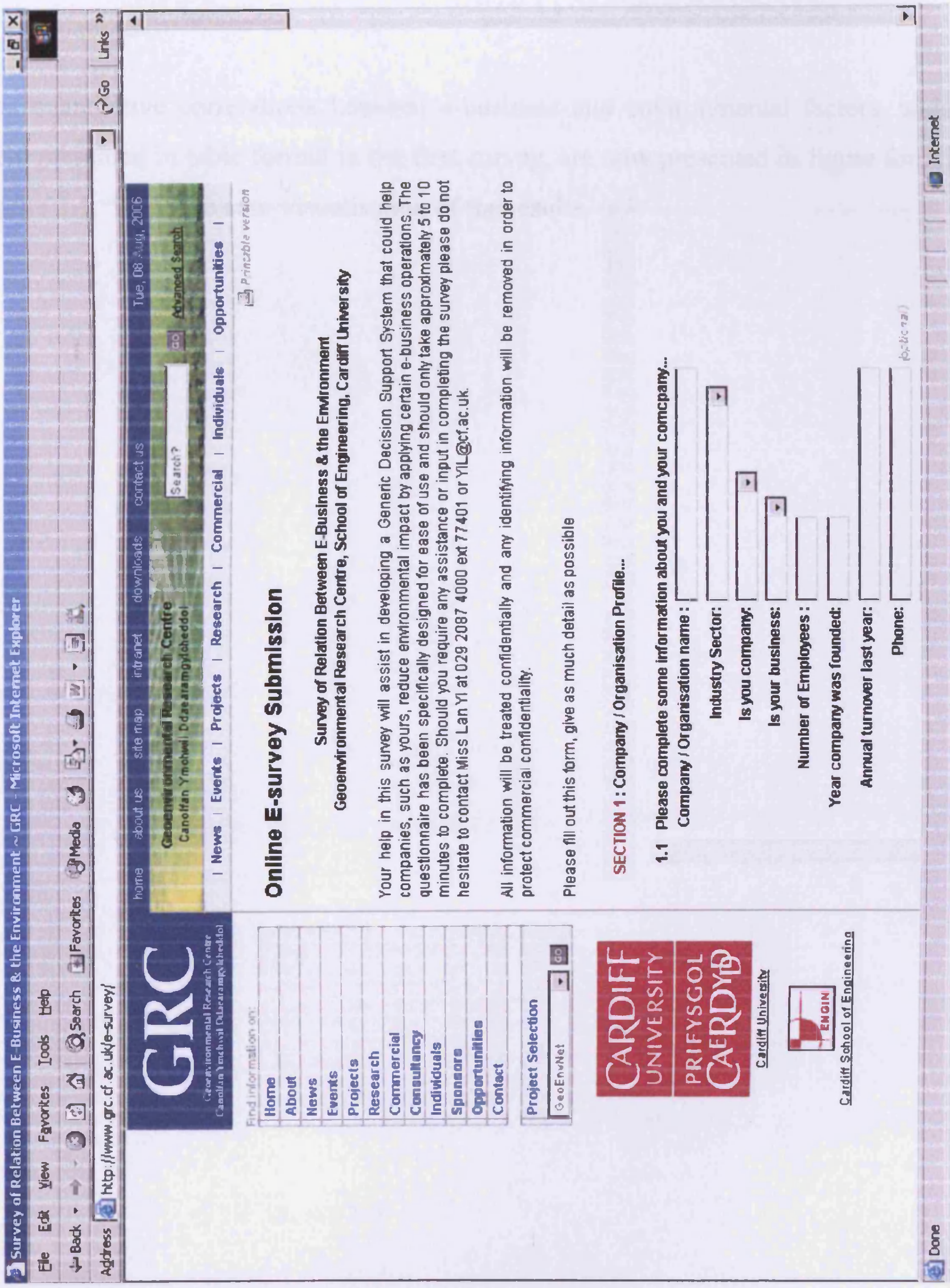



**Figure 3-2 Online survey**

## 3.8 Data Processing of Survey 2

The data collected was stored in a database built in Microsoft Access as shown in Figure 3-3.

Also the data was standardised using the same approach as used in the first survey. A conversion table based on DEFRA guidelines was built (Figure 3-4), for convenience of calculations.

The quantitative correlations between e-business and environmental factors, which were presented in table format in the first survey, are now presented in figure form in Figure 3-5. This is to ease visualisation of the results.

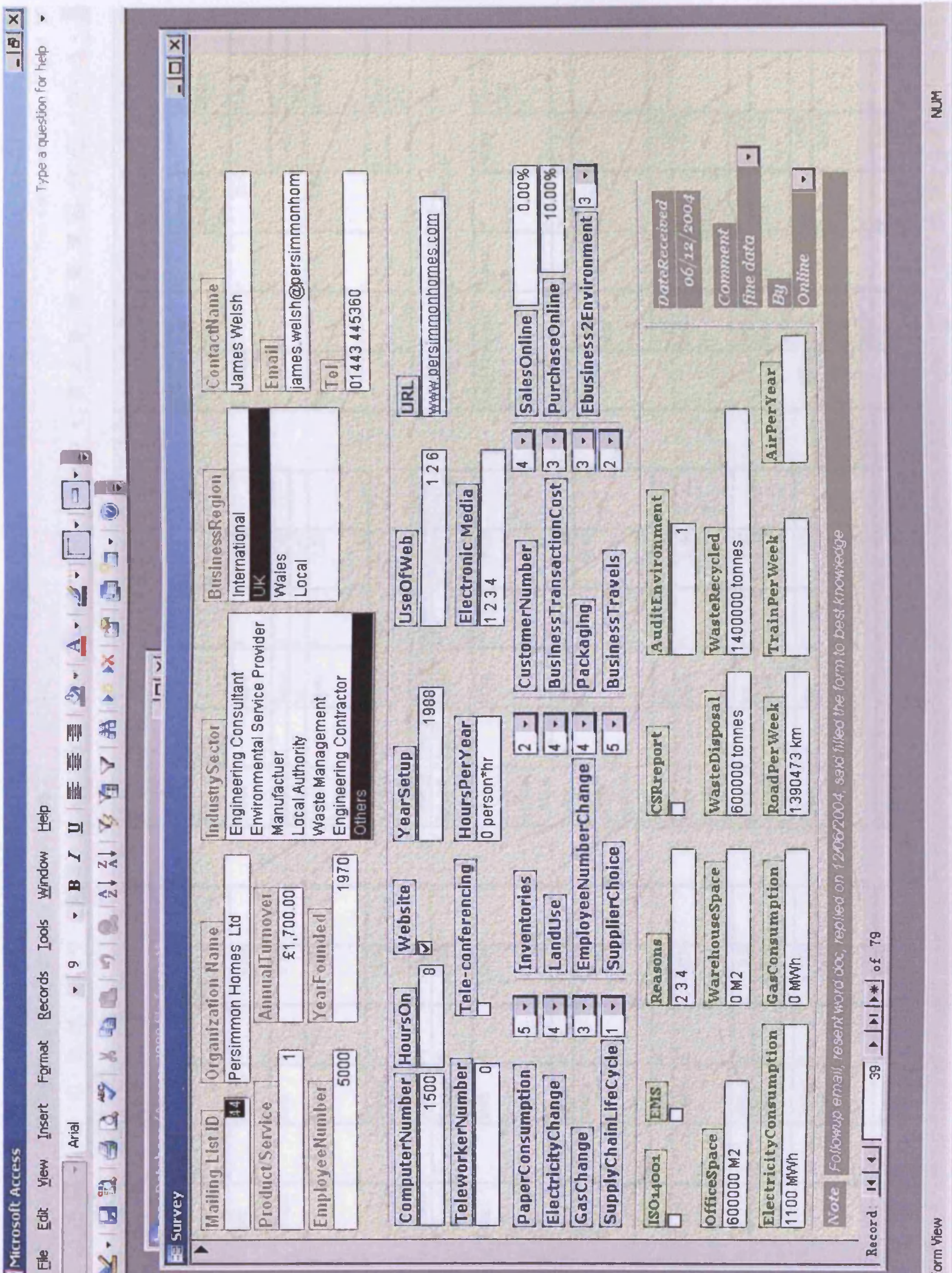
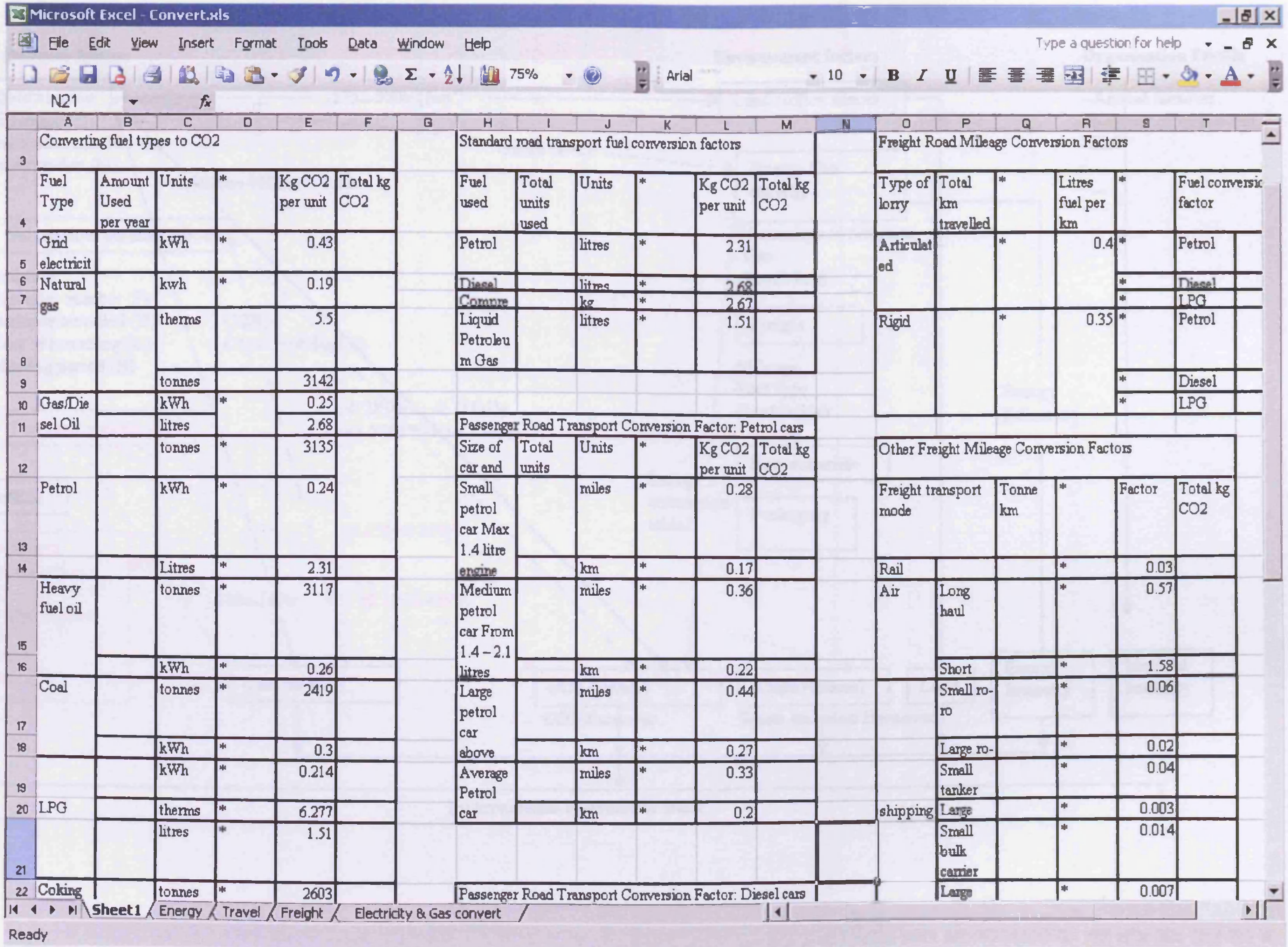



Figure 3-3 Database for survey 2




Converting fuel types to CO2

| | Fuel Type | Amount Used per year | Units | * | Kg CO2 per unit | Total kg CO2 |
|---|---|---|---|---|---|---|
| 5 | Grid electricit | | kWh | * | 0.43 | |
| 6 | Natural gas | | kwh | * | 0.19 | |
| 7/8 | | | therms | | 5.5 | |
| 9 | | | tonnes | | 3142 | |
| 10 | Gas/Diesel Oil | | kWh | * | 0.25 | |
| 11 | | | litres | | 2.68 | |
| 12 | | | tonnes | * | 3135 | |
| 13 | Petrol | | kWh | * | 0.24 | |
| 14 | | | Litres | * | 2.31 | |
| 15 | Heavy fuel oil | | tonnes | * | 3117 | |
| 16 | | | kWh | * | 0.26 | |
| 17 | Coal | | tonnes | * | 2419 | |
| 18 | | | kWh | * | 0.3 | |
| 19 | | | kWh | * | 0.214 | |
| 20 | LPG | | therms | * | 6.277 | |
| 21 | | | litres | * | 1.51 | |
| 22 | Coking | | tonnes | * | 2603 | |

Standard road transport fuel conversion factors

| Fuel used | Total units used | Units | * | Kg CO2 per unit | Total kg CO2 |
|---|---|---|---|---|---|
| Petrol | | litres | * | 2.31 | |
| Diesel | | litres | * | 2.68 | |
| Compre | | kg | * | 2.67 | |
| Liquid Petroleum Gas | | litres | * | 1.51 | |

Passenger Road Transport Conversion Factor: Petrol cars

| Size of car and | Total units | Units | * | Kg CO2 per unit | Total kg CO2 |
|---|---|---|---|---|---|
| Small petrol car Max 1.4 litre engine | | miles | * | 0.28 | |
| | | km | * | 0.17 | |
| Medium petrol car From 1.4 – 2.1 litres | | miles | * | 0.36 | |
| | | km | * | 0.22 | |
| Large petrol car above | | miles | * | 0.44 | |
| | | km | * | 0.27 | |
| Average Petrol car | | miles | * | 0.33 | |
| | | km | * | 0.2 | |

Passenger Road Transport Conversion Factor: Diesel cars

Freight Road Mileage Conversion Factors

| Type of lorry | Total km travelled | * | Litres fuel per km | * | Fuel conversio factor |
|---|---|---|---|---|---|
| Articulated | | * | 0.4 | * | Petrol |
| | | | | * | Diesel |
| | | | | * | LPG |
| Rigid | | * | 0.35 | * | Petrol |
| | | | | * | Diesel |
| | | | | * | LPG |

Other Freight Mileage Conversion Factors

| Freight transport mode | | Tonne km | * | Factor | Total kg CO2 |
|---|---|---|---|---|---|
| Rail | | | * | 0.03 | |
| Air | Long haul | | * | 0.57 | |
| | Short | | * | 1.58 | |
| | Small ro-ro | | * | 0.06 | |
| | Large ro- | | * | 0.02 | |
| | Small tanker | | * | 0.04 | |
| shipping | Large | | * | 0.003 | |
| | Small bulk carrier | | * | 0.014 | |
| | Large | | * | 0.007 | |

Figure 3-4 Conversion table

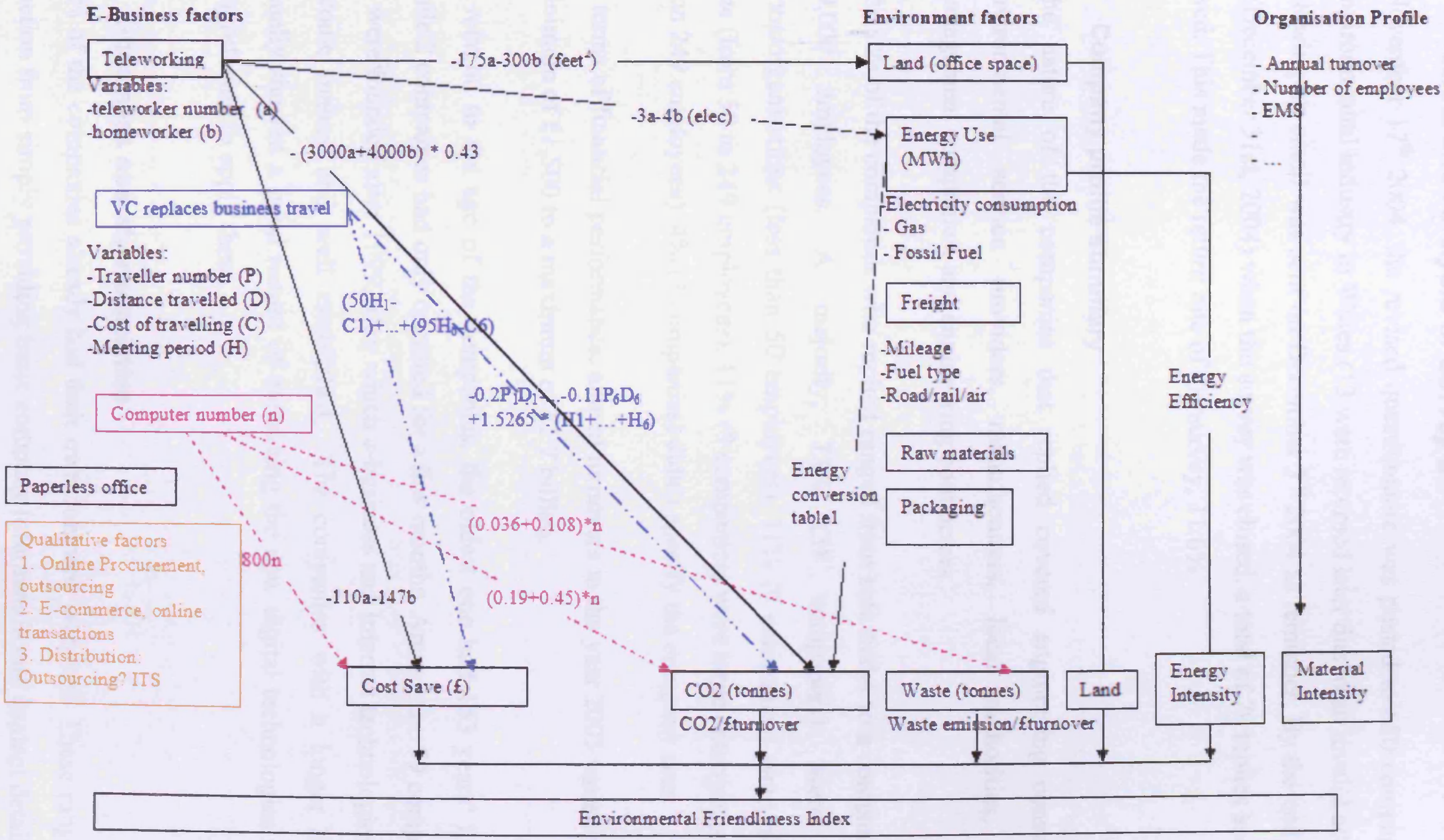



**Figure 3-5 Links between e-business factors & environmental factors**

## 3.9 Results and Analysis of Survey 2

On November 17th 2004, the revised questionnaire was posted to 680 companies in the environmental industry in Wales (13 were returned later due to an invalid address). A following up email was sent on December 3rd 2004 as reminder. By the end of the year (December 31st, 2004) when the survey was closed, a total of 79 replies had been received. This made the return rate of the survey, 11.6%.

### 3.9.1 Company profile summary

- The nature of the companies that replied covered engineering consultants, environmental service providers, manufacturers, local authorities, waste management companies, and engineering contractors.

- The size of the companies who replied ranged from sole trader to a company with 30,000 employees. A majority, 73% (58 companies), were small firms/organisations (less than 50 employees). 11% (9 companies) were medium size (from 50 to 249 employees). 11% (9 companies) were large enterprises (more than 249 employees). 4% (3 companies) didn't specify the company size.

- In terms of financial performance, annual turnovers in the year 2003 varied from a minimum of £1,500 to a maximum of £1.7 billion.

- In relation to the age of the companies, the oldest one had 153 years' history, while 2 companies had only operated for a few months. Among the 79 companies, 21 were founded after 1999, after which e-business and Internet technologies have become mature and well established. The companies with a longer history actually showed a good record of embracing the new digital technologies rather than refusing to apply them.

### 3.9.2 E-business operations/facilities

- 68% of the companies already had their own company websites. These ranged in function from simply providing basic company information and contact details for customers and partners, to providing a comprehensive e-commerce system, ie. a fully functional online transaction platform.

- The number of computers (including back-office infrastructures) ranged from 0 to 9000, which made from 0 to a maximum of 4 computers per employee. On average, 10% of the companies powered their computer for less than 5 hours per day. A majority of 60% powered for 6 to 9 hours per day. 28% powered for more than 9 hours per day. Electricity consumption of computers wasn't thought to affect companies' overall electricity consumption, as 62% companies replied their electricity consumption had not changed much since they started using Internet and related technologies to conduct business. A total of 34% replied it had decreased or increased just slightly.

- In relation to tele-working, more than 35% of the companies had either part time or fulltime teleworkers. These individuals remotely worked from home with a company Intranet, email etc rather than commuting to office. For those companies that adopted more or less teleworking, the average percentage of the employee total that were teleworkers was around 36%.

- In relation to tele-conferencing, more than 29% of the companies had replaced some travelling with some degree of tele-conferencing, from 6 hours to 8000 hours per year.

- In relation to how e-business and related technologies affect a company's day-to-day operations, results are shown in the table below.

| | Decreased a lot | Decreased slightly | Not much change | Increased slightly | Increased a lot | *Total* |
|---|---|---|---|---|---|---|
| Paper consumption | 21.52% | 30.38% | 22.78% | 15.19% | 7.59% | *97.47%* |
| Electricity consumption | 1.27% | 11.39% | 62.03% | 22.78% | 0.00% | *97.47%* |
| Gas consumption | 2.53% | 8.86% | 83.54% | 0.00% | 0.00% | *94.94%* |
| Supply chain life cycle | 3.80% | 17.72% | 67.09% | 3.80% | 1.27% | *93.67%* |
| Inventories | 1.27% | 17.72% | 65.82% | 10.13% | 0.00% | *94.94%* |
| Land use | 2.53% | 15.19% | 72.15% | 6.33% | 0.00% | *96.20%* |
| Employee number | 3.80% | 12.66% | 62.03% | 15.19% | 2.53% | *96.20%* |
| Choice of suppliers | 3.80% | 11.39% | 44.30% | 27.85% | 8.86% | *96.20%* |
| Customer number | 3.80% | 11.39% | 50.63% | 26.58% | 3.80% | *96.20%* |
| Business transaction | 8.86% | 29.11% | 51.90% | 5.06% | 1.27% | *96.20%* |
| Packaging | 3.80% | 21.52% | 63.29% | 7.59% | 0.00% | *96.20%* |
| Business travels | 2.53% | 30.38% | 46.84% | 11.39% | 3.80% | *94.94%* |
| *Average* | *4.96%* | *18.14%* | *57.70%* | *12.66%* | *2.43%* | *95.89%* |

**Table 3-8 E-business impacts from survey**

The results presented in Table 3-8 are explained in further detail below:

- Paper consumption.

Even though e-business hadn't created a completely paperless office, it had reduced the paper use for more than half the companies, with the widespread use of computers, digitisation of business transactions, email contact instead of traditional hard copy, etc. On the other hand there were still some companies who noticed their paper consumption had on the contrary increased, due to staff printing emails, internet shopping list etc. This was expected to improve in the future, by increasing employee's environmental awareness through staff training.

- Supply chain life cycle, inventory, land use and business transaction costs change.

The e-SCM (supply chain management) was expected to speed up more accurate forecasting and planning of market needs, and also more transparent and efficient communication upstream to suppliers which could avoid overproduction, improve inventory and warehouse management, and increase productivity. Among the 79 companies that participated in this survey, 22% shortened their supply chain life cycle, 19% achieved inventory reduction, 18% decreased land use, and 38% of the companies' business transaction costs were cut.

- Choices of suppliers and customers.

Empowered by the Internet and World Wide Web, companies were facing a more transparent market, with more choices of suppliers across the geographical border. 37% of the companies already took advantage of the new opportunities. On the downstream of the supply chain, customers were given more choice regarding their suppliers. Companies could gain more customers because of more accessibility to the market, or could lose their customers because of more and stronger competition. It was therefore not surprising that 15% of the companies had noticed customer numbers decreasing while 30% had noticed an increase.

- Company size change.

How did e-business affect human resources? Did it increase employee numbers because more IT staff was needed? Or did the electronic network and computers replace some chores used to be done previously by personnel? According to this survey, 16% of the companies found their employee number declined while 17% had increased their employee numbers.

- Packaging.

Unlike a traditional supply chain where manufacturers pack and ship large quantities of products to distribution centres and then on to retailers, where products are finally unpacked and placed on shelves for customers to choose, products bought online were mostly packed and delivered individually to customers. Thus there was a concern that e-business would increase packaging materials, shipping cost and corresponding waste. However the majority, 63% of the companies, reported not much change of packaging since the adoption of e-business. 25% noticed a decrease and only a minority, 8%, found a slight increase.

- Business travel.

Virtual meetings replacing physical business travel was one of the major benefits expected from e-business. 33% of the companies in this survey showed same decrease in business travel. Considering that the companies that participated in this survey were mostly SMEs in the environmental service sector, 33% is seen as quite a good performance.

- Electricity consumption.

A noticeable 23% of the companies experienced a slight increase in electricity consumption. 62% reported not much change and 11% even found a slight decrease.

- Gas consumption.

A majority, 84% of the companies, didn't have much change due to direct or indirect e-business activities.

- Online sales and purchases

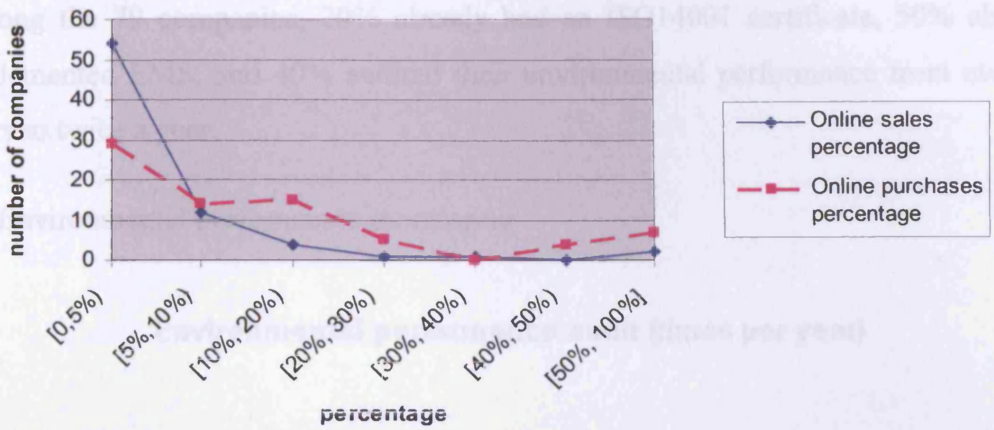



**Figure 3-6 Online sales and purchases**

- How e-business contributes to the environment

53% of the companies reported that they hadn't realised or noticed any significant change on the environment caused by e-business. However 42% of the companies noted or believed that e-business was affecting the environment in a positive way. This is quite encouraging news.

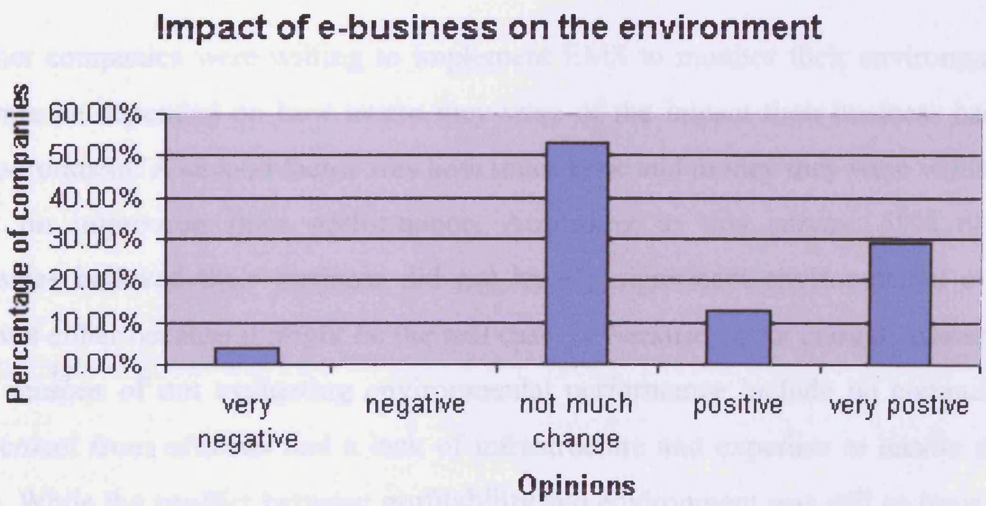



**Figure 3-7 Opinions about the impact of e-business on the environment**

### 3.9.3 Environmental performance

- ISO 14001 and EMS

Among the 79 companies, 20% already had an ISO14001 certificate, 50% already implemented EMS, and 40% audited their environmental performance from every 2 years to twice a year.

- Environmental performance monitoring

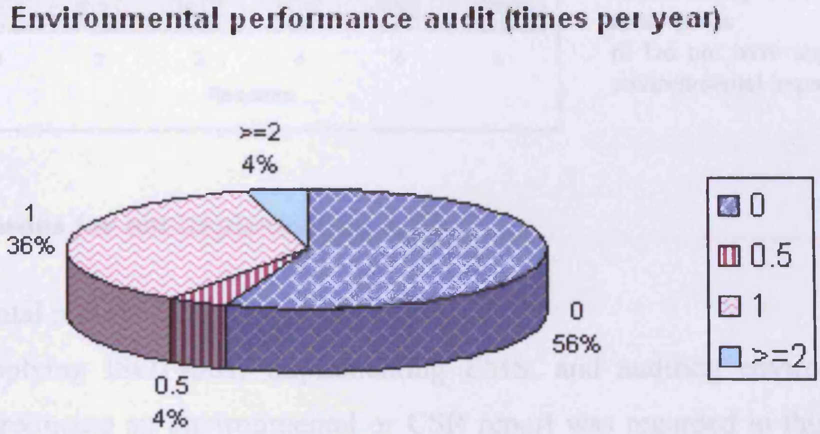
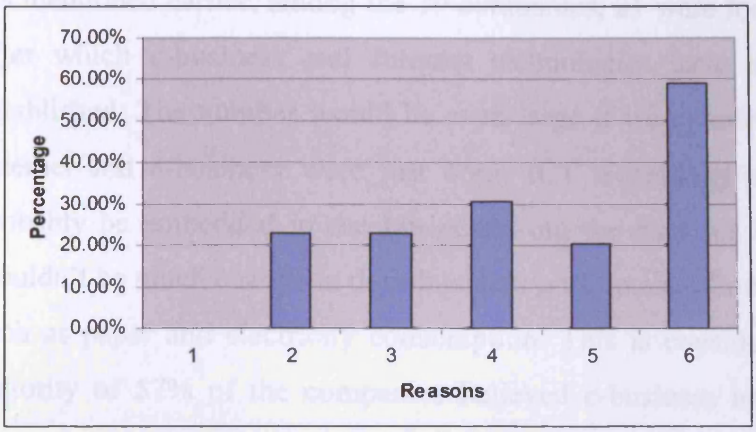



**Figure 3-8 Environment audit**

Whether companies were willing to implement EMS to monitor their environmental performance depended on how aware they were of the impact their business had on the environment. A second factor was how much time and money they were willing to invest on improving their performance. According to this survey, 59% of the companies believed their business did not have a significant environmental effect. This was either because it might be the real case, or because of not enough awareness. Other reasons of not evaluating environmental performance include no compulsory requirement from officials and a lack of infrastructure and expertise to handle these issues. While the conflict between profitability and environment was still an issue, the cost, in terms of money and time, made 51% of the companies give up monitoring environmental performance (Figure 3-9).




Reasons:
1: Simply do not care about these issues
2: No one has asked us to put such systems in place
3: No in-house expert to address these issues
4: Not enough time to think about these issues
5: Not enough money to address these issues
6: Do not have any significant environmental impact

**Figure 3-9 Reasons for not implementing EMS**

- Environmental report / CSR report

Along with applying ISO14001, implementing EMS, and auditing environmental performance, producing an environmental or CSR report was regarded in this survey as another way to measure whether a company cared about the environment. This probably required more commitment and transparency to the public than the former activities. Around 25% of the companies had started doing so, which is a good start.

- Other factors

In order to verify the e-business fantasy that "The Internet can turn buildings into Web sites and replace warehouses with supply-chain software", detailed numerical data about a company's office space and warehouse space were requested.

In order to investigate whether WEEE (waste electric and electronic equipments) has become one of the main streams of waste, information about waste disposed and recycled per year was also requested.

Road, train and air travelled for business purposes were inquired to find out whether teleconferencing and teleworking have reduced physical travels.

### 3.9.4 Other Issues

Besides the answers directly provided by the surveyed companies, in the author's opinion some other issues should also be taken into account when analysing the results.

As mentioned earlier, among the 79 companies, 21 were founded after 1999, the date after which e-business and Internet technologies have become mature and well established. The number would be even large if we counted from the date when the Internet and e-business were just born. ICT technology and facilities would most probably be embedded in the business from the start for these new-borns, so there wouldn't be much change in their business performance from traditional to e-business, such as paper and electricity consumption. This is considered as one reason why a majority of 57% of the companies believed e-business hasn't made much change. Answers regarding comparisons from these companies should be given less weight.

In this second survey, the questionnaire was sent to Welsh companies or branches in Wales only. According to the chart below, in terms of penetration of e-commerce in companies on average, Wales was lagging behind all the other regions in the UK, except Northern Ireland. Furthermore, the performance of the UK fell behind other major European countries as well as US, Canada and Australia. However the discussion, statistic data, study and case studies around this topic were mainly from these other nations. So compared with most literature reviews in this field, the result of this survey can be considered to be more conservative. E-business hasn't shown that much impact on the environment for many companies, or just not yet.

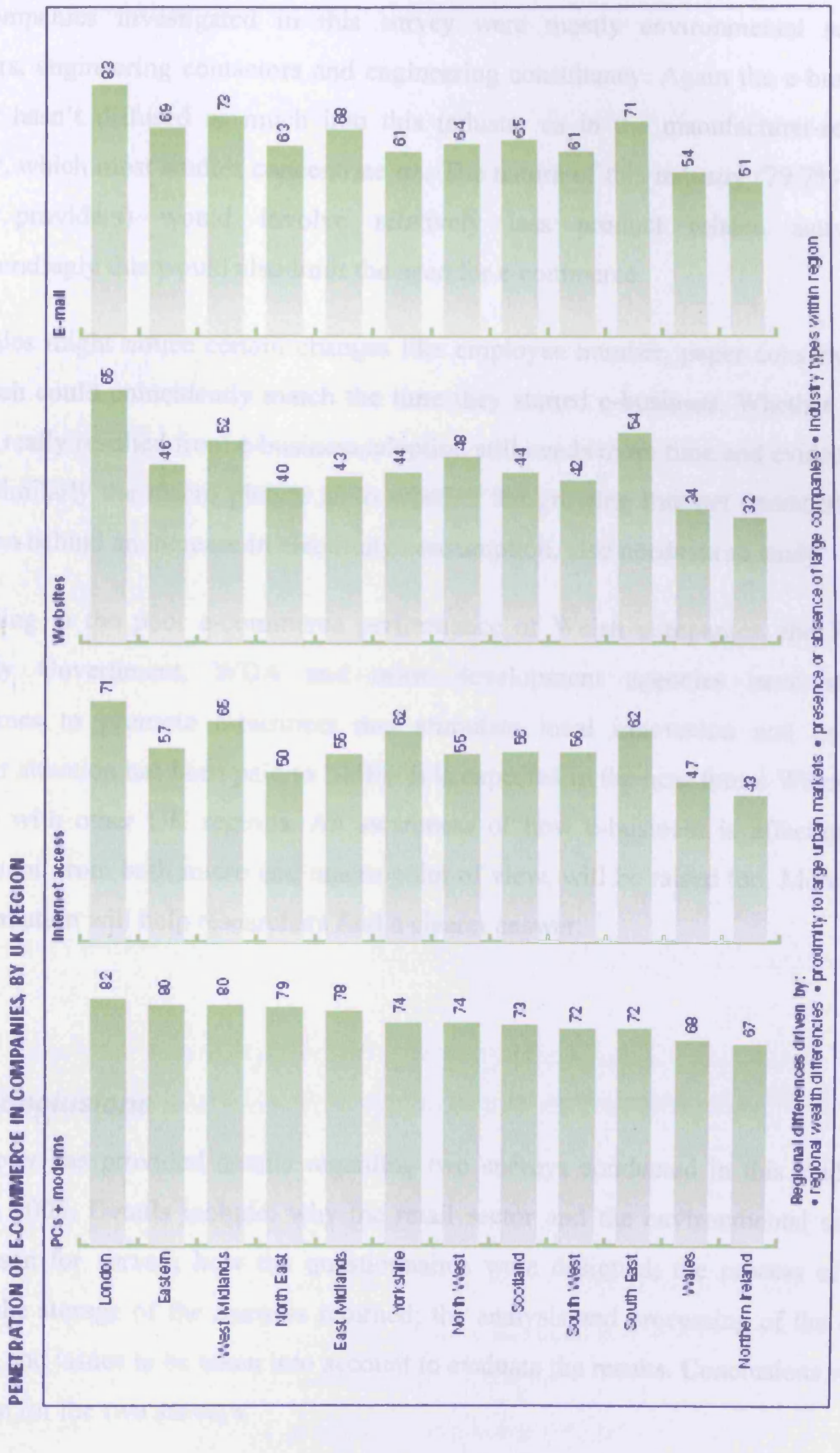



**Figure 3-10 Penetration of e-commerce in companies, by UK region (DTI 1999)**

The companies investigated in this survey were mostly environmental service providers, engineering contactors and engineering consultancy. Again the e-business concept hasn't diffused as much into this industry as in the manufacturer-retailer industry, which most studies concentrate on. The nature of this industry (79.7% were service providers) would involve relatively less product related activities. Correspondingly this would also limit the need for e-commerce.

Companies might notice certain changes like employee number, paper consumption etc, which could coincidently match the time they started e-business. Whether these changes really resulted from e-business adoption still needs more time and evidence to prove. Similarly the macro picture as to whether the growing Internet operation was the reason behind an increase in electricity consumption, also needs more study.

Responding to the poor e-commerce performance of Welsh companies, the Welsh Assembly Government, WDA and other development agencies have started programmes to promote e-business and stimulate local innovation and growth. Particular attention has been paid to SMEs. It is expected in the near future Wales will catch up with other UK regions. An awareness of how e-business is affecting the environment, from both micro and macro point of view, will be raised too. More data and information will help researchers find a clearer answer.

## 3.10 Conclusions

This chapter has provided details regarding two surveys conducted in this study in 2004 and 2005. Details include: why the retail sector and the environmental sector were chosen for survey; how the questionnaires were designed; the process of the surveys; the storage of the answers returned; the analysis and processing of the data collected; and issues to be taken into account to evaluate the results. Conclusions were also drawn for the two surveys.

The author found that data and awareness in this field were scarce, and our understanding of both the problem itself and the methodology to study it were limited. The very few guidelines available were in different structures and caused confusion for industry to produce environmental reports or Corporate Social Responsibility reports, which contains some data needed. For companies who have started measurements, results were also in various forms: absolute figures versus relative figures, quantitative methods versus qualitative methods, different units... Therefore it is critical for academics and governments to standardise the indicators and methodologies and provide best practice for industries.

The first survey was conducted in the retail sector due to its influential role in promoting environmental sustainability as well as ICT diffusion in this sector. Whether electricity consumption changed due to use of ICT hardwares, teleworking and teleconferencing substituted physical travels, e-business caused change in paper consumption and packaging etc were investigated. However the quantity of the data collected was insufficient and too poor to build a model or run a simulation.

It was therefore decided to seek further data and that a second survey should be conducted. It was decided that a second survey would be sent to members of the LRN (GRC 2006) project within the author's research centre. It was expected that these members would be more cooperative. A web page was set up to deliver the survey, which proved to be effective.

Databases were developed to store the data. The second survey achieved better results and provided some interesting findings. For example, 42% of the companies noted or believed that e-business was affecting the environment in a positive way. This is quite encouraging news. The analysis was broken down into different aspects and discussed in detail.

The author also suggested that some other factors should be taken into account when assessing these results. For example, 27% of the companies were founded after the e-business/ICT had become mature. Therefore it is reasonable that they didn't notice any significant change caused by e-business/ICT integration. Also the second survey was limited to Wales. In terms of penetration of e-commerce in companies on average, Wales is lagging behind all the other regions in the UK.

The data and information provided the basic material needed to build a neural network model and a decision support system for this study. The survey and data collection are also considered to be a contribution to a database and knowledge base in this research area.

# Chapter 4

# Data Mining and Data Modelling

## 4.1 Introduction

The data collected from the surveys and other studies were "numbers". These included categorical data (e.g. industry sector), binary data (yes or no), discrete figures (e.g. ranking from 1 to 5), and continuous figures (e.g. energy consumption in kilowatt hour). These numbers, however, were still numerical presentations that did not yield direct information. Therefore a data mining process would be needed, a process to transform data to information and knowledge that could make academic, practical or scientific sense. These include trends, patterns, relationships and rules, with which decision-makers could make strategic decisions. The two main methodologies chosen to carry out the data-mining job in this study were fuzzy logic and neural networks.

This chapter will discuss and explain:

- Data mining in general; why it was needed and what did it do in this study.
- Fuzzy Logic (FL) and Artificial Neural Networks (ANN); why they were chosen and what did they achieve in this study.
- How the data was pre-processed using Fuzzy Logic and modelled using Artificial Neural Networks, the implementation procedure.
- An innovative approach of developing modular neural networks using correlation coefficients; why this was chosen over one overall neural network.
- The software chosen, with which the models would be built (this was mainly Matlab 7.0 Fuzzy Logic Toolbox and Neural Network Toolbox); what the

alternative software programmes were and what advantages the chosen ones provided compared with the others

- The training, testing and validation of three modular neural networks.
- The implementation of three modular neural networks for analysing how e-business/ICT indicators affect environmental indicators, which forms the engine of the forward chain simulation of the Expert Decision Support System (DSS). (The backward chain – how to change the inputs, i.e. the e-business/ICT indicators, to achieve certain environmental performance goals - will be explained in a subsequent chapter regarding knowledge discovery and rule induction)

## *4.2 Methodologies*

This section will give a brief introduction and general background of data mining, Fuzzy Logic and Neural Network, before presenting why and how they were applied and implemented in this study.

### 4.2.1 Data mining

As Han and Kamber introduced in the book entitled *Data Mining – concepts and techniques* (Han and Kamber 2001), data mining is the process of discovering interesting knowledge from large amounts of data stored either in databases, data warehouses, or other information repositories. Knowledge discovery as a process consists of an iterative sequence of the following steps:

i) Data cleaning (to remove noise and inconsistent data).
ii) Data integration (where multiple data sources might be combined).
iii) Data selection (where data relevant to the analysis task is retrieved from the database).
iv) Data transformation (where data is transformed or consolidated into forms appropriate for mining by performing summary or aggregation operations, for instance).
v) Data mining (an essential process where intelligent methods are applied in order to extract data patterns)

vi) Pattern evaluation (to identify interesting patterns representing knowledge based on some "interestingness measures")

vii) Knowledge presentation (where visualisation and knowledge representation techniques are used to present the mined knowledge to the user)

Data mining involves an integration of techniques from a multiple of disciplines such as database technology, statistics, machine learning, high-performance computing, pattern recognition, neural networks, data visualisation, information retrieval, image and signal processing, and spatial data analysis. The discovered knowledge can be applied to decision-making, process control, information management, and query processing. Therefore, data mining is considered to be one of the most important frontiers in database systems and one of the most promising interdisciplinary developments in the information industry.

### 4.2.2 Fuzzy Logic

Fuzzy Logic was initiated in by Professor Lotfi A. Zadeh at the University of California in Berkeley in 1965 (Zadeh 1965). Fuzzy logic started with the concept of a fuzzy set. The conventional set in classical mathematics is a crisp set that an object either belongs to or does not. For example, 3 is a member of the positive integer set and -4.5 is not; Saturday belongs to the weekend set and Wednesday does not; spring is a member of the season set but not a month set.

However in the real world it is not always easy to define such a crisp boundary and be precise. As Cox stated, most of the phenomena we encounter every day are imprecise; that is, they carry a certain degree of fuzziness in the description of their nature. Human beings are able to formulate plans, make decisions, and recognise compatible concepts at high levels of vagueness and ambiguity. Statements like "when engine temperature is hot, increase fan speed", "current inflation rates are rising rapidly" can be readily understood, even if we did not have an exact threshold point for the term "hot" and "rapidly" (Cox 1999).

One of the most commonly used example is a set of tall people (MathWorks 1995-2005). If the set of tall people is given the well-defined (crisp) boundary of a classical set, we might say all people taller than 6 feet are officially considered tall. So a person who is 5.999 feet is short while another one with 6.001 feet is tall. It is obvious unreasonable to call one person short and another tall when they differ in height by the width of a hair. In this case, a fuzzy set is more practical. This is a set without a crisp, clearly defined boundary and can contain elements with only a partial degree of membership. So if one asks the question "Is X a member of set A", the answer might be yes, no, or any one of a thousand intermediate values in between. Fuzzy logic is all about the relative importance of precision (Figure 4-1).

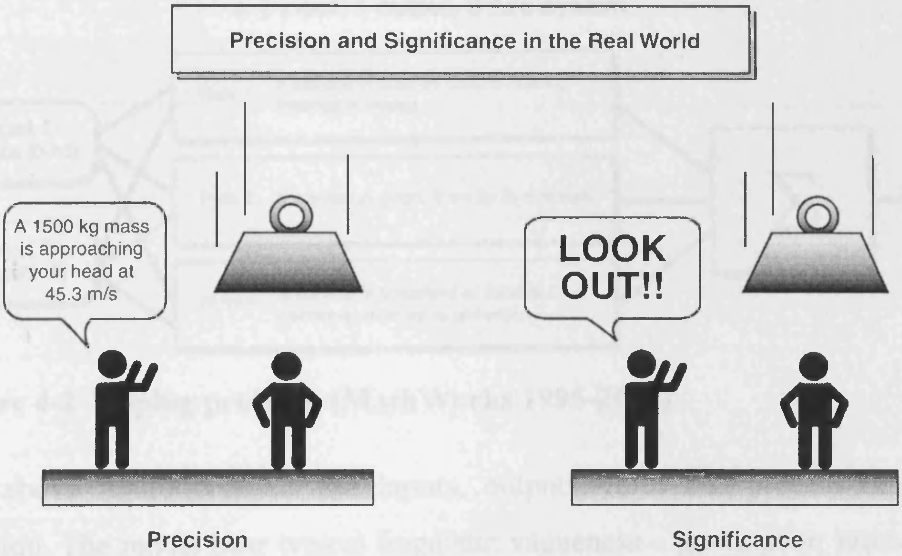



**Figure 4-1 Precision and significance (MathWorks 1995-2005)**

Fuzzy logic is derived from fuzzy set theory and an extension of Boolean logic dealing with concept of partial truth where variables can have degrees of truth or falsehood.

Fuzzy sets and fuzzy operators (logical connectives/operators: AND (intersection or conjunction), OR (union or disjunction) and NOT (complement)) are the subjects and verbs of fuzzy logic. The if-then rule statements are used to formulate the conditional statements. A membership function (MF) is a curve that defines how each point in the input space is mapped to a membership value (or degree of membership) between 0 and 1. Fuzzy inference is the process of formulating the mapping from a given input to an output using fuzzy logic.

The most common membership functions include linear functions, the Gaussian distribution function, the sigmoid curve, the quadratic and cubic polynomial curves etc. The selection is wide for those who want to explore the possibilities, but exotic membership functions are by no means required for perfectly good fuzzy inference systems (MathWorks 1995-2005).

Matlab user guide has given a very good practical example of fuzzy logic application in real life – a tipping problem in the restaurant. Given a number between 0 and 10 that represents the quality of service and food at a restaurant, what should the tip be?

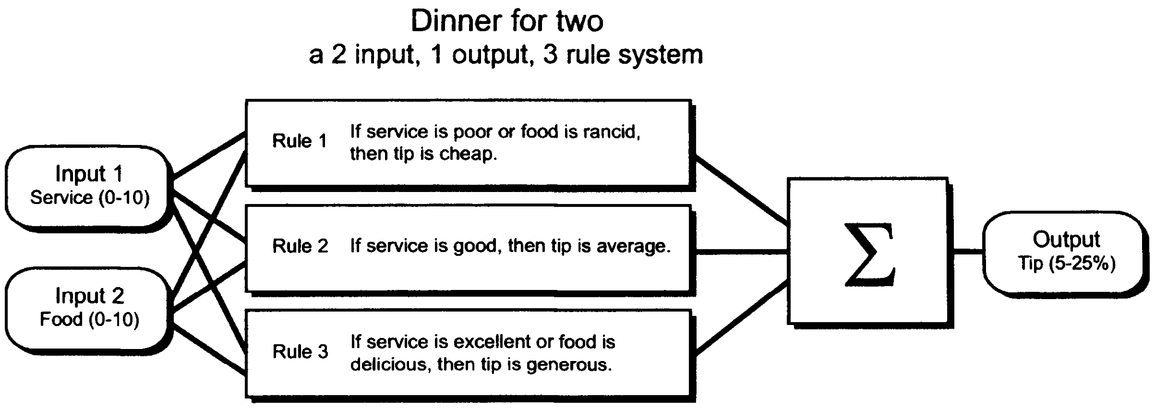



**Figure 4-2 Tipping problem (MathWorks 1995-2005)**

The above figure presents the inputs, outputs, rules and process of making the decision. The inputs bear typical linguistic vagueness – good, poor, rancid, delicious. So how does fuzzy logic solve this problem?

Figure 4-3 presents the fuzzy solution for this problem. Following the standard fuzzy computing which includes five steps: fuzzify inputs – apply fuzzy operators – apply implication method (IF-THEN rule) – aggregate all outputs (fuzzy sets that represent the outputs of each rule are combined into one single fuzzy set) – defuzzify (resolve the fuzzy set, which is a range of output values, to one single output value) (MathWorks 1995-2005), giving a 3 out of 10 rating for service and 8 for food, we got an answer – pay 16.7% tip!

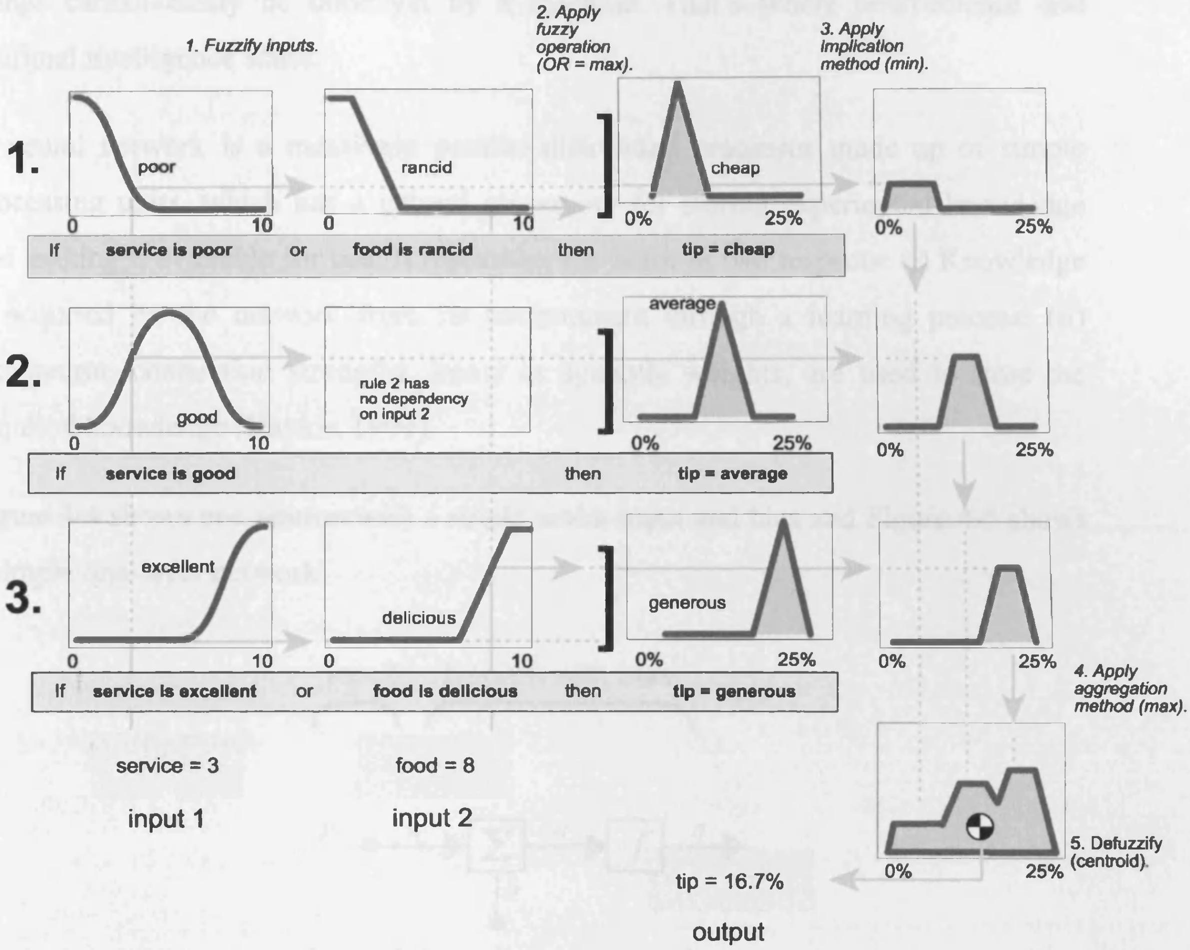



**Figure 4-3 Fuzzy inference system for tipping problem**

Of course fuzzy logic applications go far beyond the tipping problem. The past few years have witnessed a rapid growth in the number and variety of applications of fuzzy logic. The applications range from consumer products such as cameras, camcorders, washing machines, and microwave ovens to industrial process control, medical instrumentation, decision-support systems, and portfolio selection (MathWorks 1995-2005).

### 4.2.3 Artificial Neural Networks

Computers have evolved to do many things that human beings cannot handle, such as complex mathematic calculations in seconds. However the human brain still remains superior to computing machines in many aspects. For example, one can recognise a friend who one hasn't seen for thirty years; one can guess the whole picture while only seeing a partial pattern; one can read signatures from different handwritings. These things cannot easily be done yet by a machine. That's where neuroscience and artificial intelligence stand.

A neural network is a massively parallel-distributed processor made up of simple processing units, which has a natural propensity for storing experiential knowledge and making it available for use. It resembles the brain in two respects: (i) Knowledge is acquired by the network from its environment through a learning process; (ii) Interneuron connection strengths, know as synaptic weights, are used to store the acquired knowledge (Haykin 1999).

Figure 4-4 shows one neuron with a single scalar input and bias and Figure 4-5 shows a simple one-layer network.

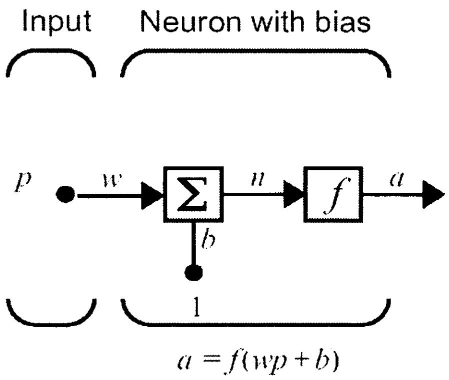



Figure 4-4 A neuron with a single scalar input and bias (Demuth and Beale 1992-2004)

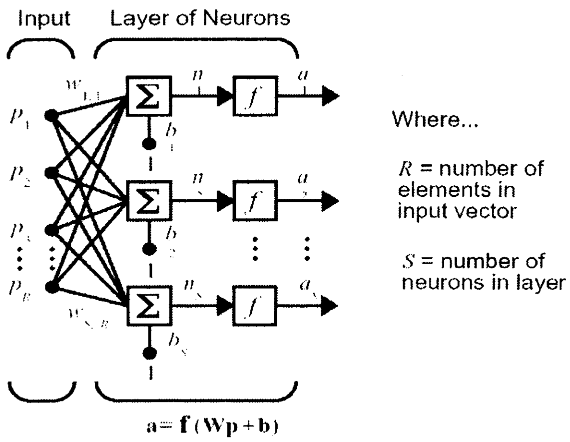



Figure 4-5 A one-layer network with R input elements and S neurons (Demuth and Beale 1992-2004)

The basic components and concepts of a neural network include (Nelson and Illingworth 1991):

- Neuron or processing element (PE) – to evaluate the input signals, determine the strength of each one, calculate a total for the combined input signals and compare that total to some threshold level, and determine what the output should be.
- Inputs and outputs – all of input signals come into PE simultaneously. A neuron either 'fires' or 'doesn't fire' depending on some threshold level.
- Bias – just as real neurons are affected by things other than inputs, some networks provide a mechanism for other influences. Sometimes this extra input is called a bias term, or a forcing term.
- Weighting factors – adaptive coefficients within the network that determine the intensity of the input signal. Initial weights may be set by an algorithm or set at random. How the interconnection weights are changed is the function of the *learning algorithm.*
- Summation functions - all of the products are summed and compared to some threshold to determine the output. If the sum of the inputs is greater than the threshold value, the processing element generates a signal; otherwise no signal (or some inhibitor signal) is generated.
- Activation Functions - the result of the summation function could be input to an activation function before being passed on to the transfer function.

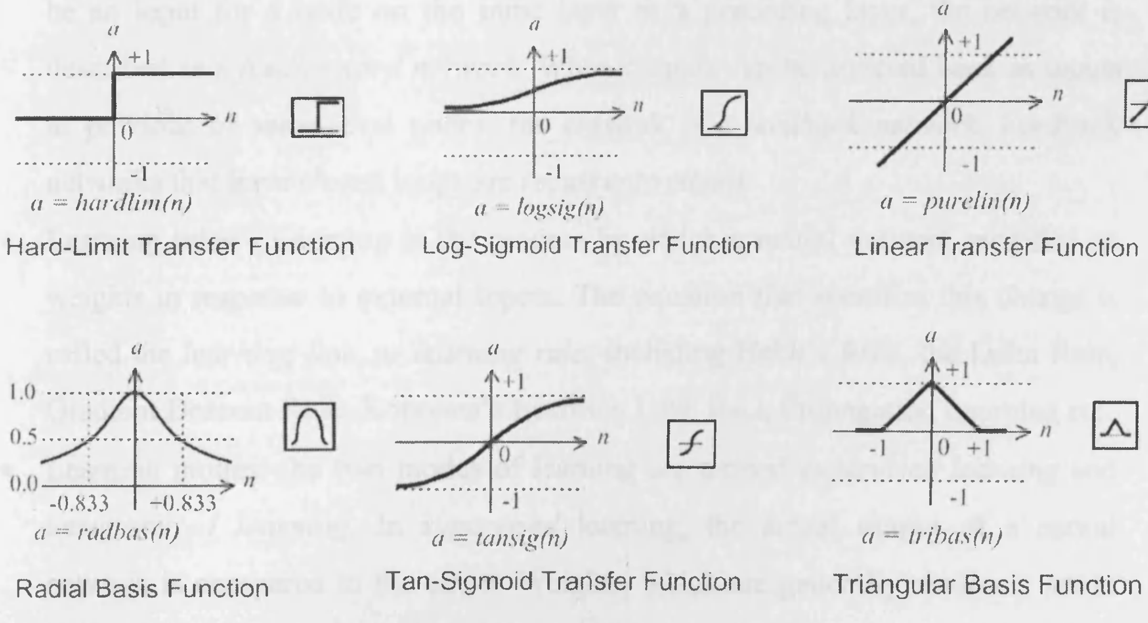



**Figure 4-6 Sample transfer functions (Demuth and Beale 1992-2004)**

- Transfer Functions - the threshold to decide whether a signal is generated from PE, it is generally nonlinear (Figure 4-6 shows common transfer functions).
- Learning functions – input signals can be positive (excitatory) or negative (inhibitory). A positive input promotes the firing of the PE, whereas a negative input tends to keep the PE firing. If we attach some local memory to our PE, we can store results of previous computations and modify the weights used as we go along. This ability to change the weights allows the PE to modify its behaviour in response to its inputs, or *learn*.
- Combining layers – the layer that receives the inputs is called the *input layer*, which typically performs no function other than the buffering of the input signal. The network outputs are generated from the *output layer*, which makes the neural network information available to the outside world. Any other layers are called *hidden layers*. Sometimes they are likened to a "black box" within the network system. Whenever the structure of input pattern is quite different from that of the output pattern, hidden units are needed to create an internal representation from the input signals. The ability to process information increases in proportion to the number of layers in the network.
- Connectivity options – connectivity has to do with how the outputs are channelled to become inputs. The output signal from a node may be passed on as input to other PE, or even possibly sent back as an input to itself. When no PE output can be an input for a node on the same layer or a preceding layer, the network is described as a *feedforward network*. When outputs can be directed back as inputs to previous or same-level nodes, the network is a *feedback network*. Feedback networks that have closed loops are *recurrent systems*.
- Learning rules - Learning is the process by which a neural network modifies its weights in response to external inputs. The equation that specifies this change is called the *learning law*, or *learning rule*, including Hebb's Rule, the Delta Rule, Gradient Descent Rule, Kohonen's Learning Law, Back Propagation Learning etc.
- Learning modes: the two modes of learning are termed *supervised learning* and *unsupervised learning*. In *supervised* learning, the actual output of a neural network is compared to the target. Weights, which are generally randomly set to begin with, are then adjusted by the network so that the next iteration will produce

a closer math. The goal of all learning procedures is ultimately to minimise the error between the target and the current output sample by continuously modifying the weights. In *unsupervised* learning, networks use no external influences to adjust their weights. Instead there is an internal monitoring of performance. The network looks for regularities or trends in the input signals, and makes adaptations according to the function of the network.

Figure 4-7 shows an example of a multiple layer neural network architecture.

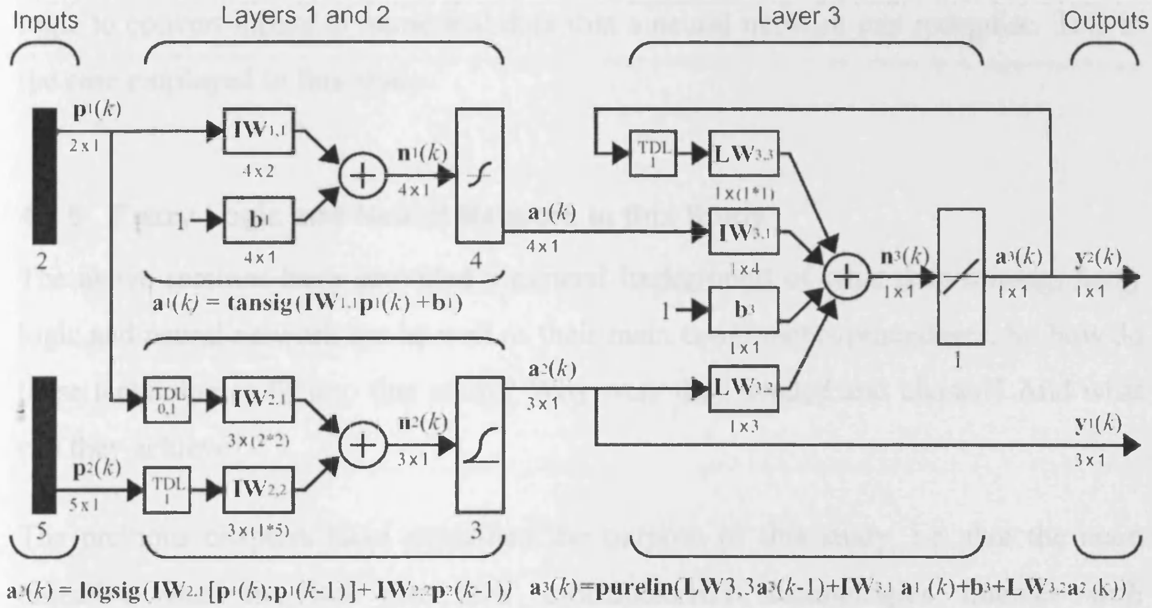



**Figure 4-7 A multiple layer neural network (Demuth and Beale 1992-2004)**

Typical neural network paradigms/models include: perceptron, ADALINE/MADALINE, Brain-State-in-a-Box (BSB), Hopfield network, Back Propagation, Self-Organising Maps.

### 4.2.4 Fuzzy Neural Net

The combination of fuzzy logic and neural network has a twofold application.

First, the *adaptive neuro-fuzzy inference system* (ANFIS), the membership function parameters of a fuzzy inference system (FIS) are tuned/adjusted using neural network technologies such as the backpropagation algorithm alone, or in combination with a least squares type of methods. This technique provides a method for the fuzzy modelling procedure to learn information about a data set, in order to compute the membership function parameters that best allow the associated FIS to track the given input/output data (MathWorks 1995-2005). ANFIS has recently been developed for business and industrial applications, for example for assessing the risk of fraudulent financial reporting (Lin et al. 2003), and traffic control (Henry et al. 1998).

The second application is where the fuzzy neural combination is not about embedding neural technologies in a fuzzy system for a membership function, but using fuzzy logic to convert inputs to numerical data that a neural network can recognise. This is the case employed in this study.

### 4.2.5 Fuzzy Logic and Neural Network in this Study

The above sections have provided a general background of what data mining, fuzzy logic and neural network are as well as their main components/procedures. So how do these technologies fit into this study? Why were they needed and chosen? And what can they achieve?

The previous chapters have explained the purpose of this study, i.e. that the main objective was to find out how e-business/ICT technologies interact with environmental performance within a certain industry/sector. Indicators were designed/chosen to represent the main features of the two sides (e-business and environmental) respectively, based on a literature review and interviews with industry experts. Questions were asked and answers received around these indicators in two surveys. Now the data was ready, what next? Values of e-business/ICT attributes $A_1$, $A_2$, $A_3$ … $A_m$ are available on one side and values of the environmental performance indicators $B_1$, $B_2$, $B_3$ … $B_n$ on the other. However the way in which each indicator on one side is connected to indicators on the other side is not known. The question is: is $A_1$ connected to $B_3$ or not? If so, what is the weight of this relationship? For example, $A_1$ might make a 10% contribution to $B_3$ while $A_2$ might make a 50% contribution. Also is the relationship linear, that is

$$B_1 = A_1 * W_1\% + A_2 * W_2\% + \ldots + A_m * W_m\% \qquad (2)$$

or is it much more complicated? The study has presented a typical black-box problem. So what is the solution for this kind of problem?

After studies and comparisons, a Neural Network approach stood out to be the closest candidate. The main reasons are as follows:

- The approach deals with poor quality or incomplete data with a strong fault tolerance, which is the exact case for the survey data collected.
- The approach works with a situation where it is difficult to specify a model for mathematical simulation, or rules for a knowledge based system.
- The approach is capable of learning from examples available through training processes.
- The approach's generalisation capability is very useful, for predicting answers for unseen situations.
- The approach is adaptive – when more training data are available, the new data can be applied to the existing neural network to improve its performance. This is very helpful and promising for this research topic.
- The approach can handle multi-variable, multi-dimensional space but requiring less computational demand/efficiency than other methods.
- The approach's typical application includes analysis of databases to determine patterns inherent in data.
- Not only are neural networks structurally parallel, but also the processing sequence is parallel and simultaneous, so very fast decisions will be possible and will be able to be made in real time. (Nelson and Illingworth 1991)
- The approach can work as a component within an overall application (the DSS in this case) and can be used in conjunction with both conventional processing techniques (the programming of DSS in this case) and rule-based processing (the rule induction and backward chain reasoning of the DSS).

Other technologies applied in this research field by previous studies and reviewed in the Methodologies section of the Literature Review chapter include scenarios, LCA and Ecological Footprint. Generic data mining technologies have also been applied, including prediction methods (linear and multiple regression, logistic regression) and classification methods (Decision tree, Bayesian Classification, k-nearest neighbour, Genetic Algorithm, rough set, fuzzy set etc) (Han and Kamber 2001). However none can offer all advantages available from a neural network.

But, availability of data is an essential requirement for neural nets. Neural networks require large quantities of data at every stage of development. A supervised neural network requires both input data, which constitutes the problem, and target data (DTI 1994). So the data is critical for the success of a neural network prediction.

Also neural networks deal only with numeric input data. Therefore, one must often convert or encode data from the external environment (Nelson and Illingworth 1991). The data collected in the survey were not all numerical, e.g. had paper consumption increased or decreased; was the e-business/ICT contribution to the environmental performance rated very negatively to very positively. This linguistic vagueness of 'increase' and 'positive' etc cannot be recognised by the neural network. In order to transform these terms to something a neural network can process, a conversion would be needed, which is exactly what fuzzy set theory can offer.

Similar to the tipping problem, suppose a company had a significant increase in electricity consumption, not much change in gas consumption, and a slight decrease in business travel, does that mean anything to the environment? Maybe they all contribute to $CO_2$ emissions? But then what is the effect? 25% increase? That's where fuzzy sets and fuzzy logic stand in this study. Considering a neural network only takes numerical data, a fuzzy logic is able to convert the linguistic attributes to numbers a neural network can recognise. On the other hand, large numbers of inputs would need huge amount of data and a longer process to train. Fuzzy logic also helps to reduce the number of attributes and training cycles. Another advantage of fuzzy logic is, in the neural network development environment (Matlab from Mathworks), a fuzzy logic toolbox can also be added on, which can smooth the integration with the neural network implementation.

## 4.3 Software

First of all, it should be pointed out that programming a neural network is different to conventional programming. Nelson and Illingworth (Nelson and Illingworth 1991) stated the job of neural net programming as follows: the programmer would not write algorithms but would need to specify transfer functions, training laws, and the structure of the network. In addition to all of this, some sort of scheduling function would be needed to determine if and how often processing elements would update (continuously or periodically). This new breed of programmer would need to understand statistics in order to select training sets and evaluate output result adequately. Neural networks require a different set of skills from that required by conventional programming. Also, the focus of this study was to analyse the relationship of e-business/ICT indicators and environmental performance indicators with the help of neural network technology, rather than carry out research in a neural network itself. Writing a code for a neural network from scratch wasn't therefore necessary.

In order to develop the neural network, the environment within which it would be developed it had to be chosen first. Searching neural network tools or software in google.com, around 15,700,000 results would appear (20/09/2006), linking to sites and programs developed by various companies or individuals with an emphasis on various types of neural networks for different purposes. The results include:

- Commercial software: computer software sold for commercial purposes or that serves commercial purposes

- Freeware: computer software which is made available free of charge, although typically freeware is proprietary. It is usually distributed without source code and often carries a restrictive license. For example, a license might allow the software to be freely copied, but not sold, or might forbid commercial, government or military use

- Shareware: a marketing method for commercial software, whereby a trial version is distributed in advance and without payment, as is common for proprietary software

- Open source software, which is computer software whose source code is available under a copyright license that permits users to study, change, and improve the software, and to redistribute it in modified or unmodified form.

Among these software programmes, some are standalone applications, some are components of bigger software programmes, and some are to be integrated into other software such as Microsoft Excel. A few of the software programmes were tested. However, for some, the user instructions were obscure and the software was not well documented. Reliability was thus in doubt.

After some research and comparison, the choices were finally limited to three software programmes: Weka, Clementine and Matlab.

### 4.3.1 Weka

Weka is a collection of machine learning algorithms for data mining tasks developed by the University of Waikato. The algorithms can either be applied directly to a dataset or called from the developer's own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also suited for the development of new machine learning schemes. It is open source software, issued under GNU General Public License. Weka 3.4 was the latest version and can be downloaded from site http://www.cs.waikato.ac.nz/ml/weka/.

Figure 4-8 presents the general environment of Weka (GUI, Explorer, Knowledge flow environment, Experiment environment).

In the author's opinion, Weka is a comprehensive data mining software programme with the modules corresponding to the tasks in data mining such as pre-processing, classification, cluster and so on. But it is not strong in neural network implementation compared with Matlab. It is also not strong in rule induction and knowledge discovery compared with the Clementine. These facts will be explained in detail below.


**Figure 4-8 Weka 3.4**

### 4.3.2 Clementine

Clementine® is an enterprise data mining software programme from SPSS Company, who developed the well known SPSS statistical software. It was first developed by Integral Solutions Ltd in UK. It then became SPSS's data mining workbench after the company was acquired by SPSS in 1998. Earlier versions could still be branded under Integral Solutions.

Clementine offers visual interactive workflow interface and predictive modelling to support the data mining process. It is designed for ease of use so that data miners can focus on knowledge discovery rather than on technical tasks like writing codes. Also it claims to be able to perform in-database modelling in major database brands such as IBM®, Oracle® and Microsoft® (SPSS 2005).

In the author's opinion, Clementine is the easiest to use and the visualisation is the best among the three candidate software programmes for this study. Users can simply drag and drop objects (streams, graphs, models) in the working environment without understanding embedded algorithms and functions within it. However the availability of algorithms and functions are relatively limited in the neural network part. It was therefore chosen to fulfil the rule induction and knowledge discovery – the backward chain of the DSS in this study.

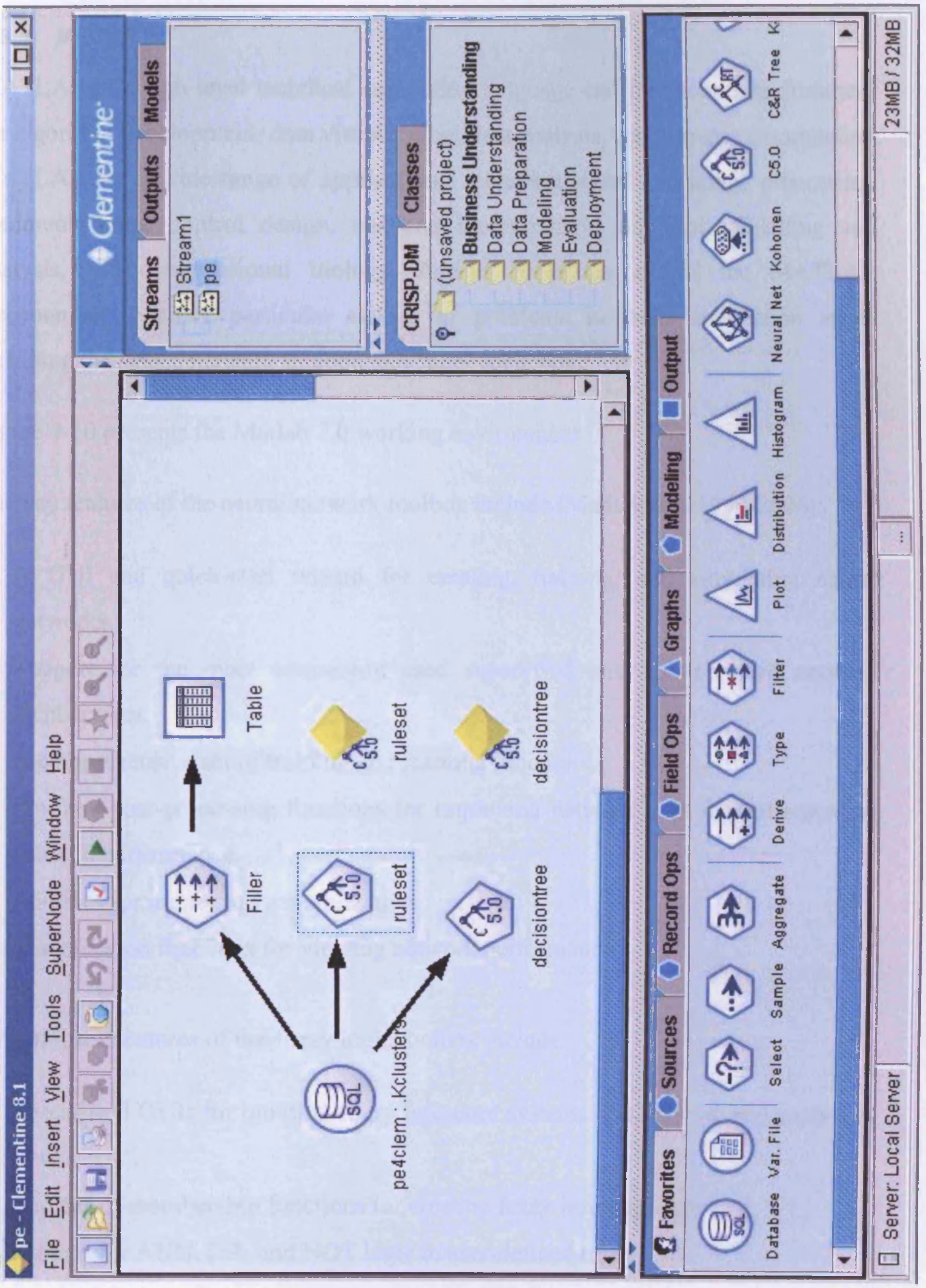



**Figure 4-9 Clementine 8.1**

### 4.3.3 Matlab

MATLAB is a high-level technical computing language and interactive environment for algorithm development, data visualization, data analysis, and numeric computation. MATLAB has a wide range of applications, including signal and image processing, communications, control design, test and measurement, financial modelling and analysis, and computational biology. Add-on toolboxes extend the MATLAB environment to solve particular classes of problems in these application areas, including the neural network toolbox and fuzzy logic box.

Figure 4-10 presents the Matlab 7.0 working environment.

The key features of the neural network toolbox include (MathWorks 1994-2006):

- A GUI and quick-start wizard for creating, training, and simulating neural networks.
- Support for the most commonly used supervised and unsupervised network architectures.
- A comprehensive set of training and learning functions.
- Pre- and post-processing functions for improving network training and assessing network performance.
- Routines for improving generalization.
- Visualization functions for viewing network performance.

Also, the key features of the fuzzy logic toolbox include:

- Specialized GUIs for building fuzzy inference systems and viewing and analysing results.
- Choices of membership functions for creating fuzzy inference systems.
- Support for AND, OR, and NOT logic in user-defined rules.
- Standard fuzzy inference systems.

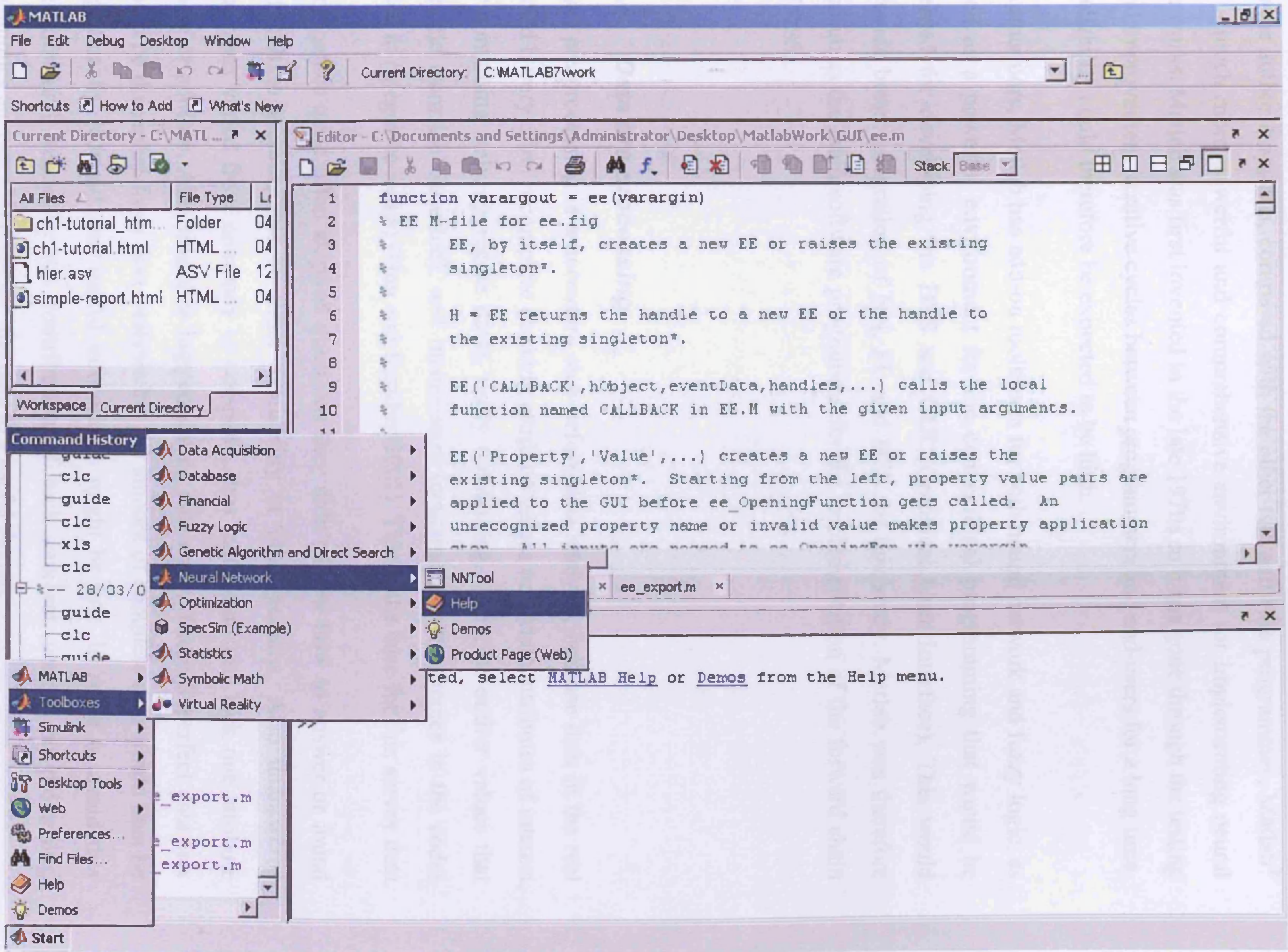



Figure 4-10 Matlab 7.0

In the author's opinion, compared with the other two software programmes, Matlab® is a much more powerful and comprehensive environment for implementing neural networks. Matlab was first invented in the late 1970s so it has gone through the testing and improvement iterative cycles between programmers and end-users for a long time. Reliability could therefore be expected to be high.

Furthermore, Matlab has add-on toolboxes for both neural network and fuzzy logic, as well as a powerful environment for the conventional programming that would be needed for developing the DSS and GUI (Graphical User Interface). This would provide better integration of NN, FL and DSS for this study. Matlab was therefore chosen as the main software programme for the implementation of the forward chain of DSS.

## *4.4 Data pre-processing*

Data pre-processing is a necessary step before data analysis, because data in the real world is very often incomplete (lacking attribute values or certain attributes of interest, or containing only aggregate data), noisy (containing errors, or outlier values that deviate from the expected), and inconsistent (containing discrepancies in the codes used to categorise items) (Han and Kamber 2001). This is the case for this survey data.

Companies quite often skipped questions they didn't know how to answer or found difficult to answer, due to time, availability or other reasons. Also unanswered questions varied from company to company, thus many data sets lack one attribute value or another. Also mistakes happen during data entry. These imperfect data sets could be discarded from data analysis but the amount of complete data would thus be significantly reduced and useful information might be lost. In order to avoid this situation and maximise the contribution of each data set, certain pre-processing techniques are employed to 'fix' the data to a useful degree.

### 4.4.1 Data Cleaning

- Outliers

Outliers are data objects that do not comply with the general behaviour or model of the data. Outliers may be detected using statistical tests that assume a distribution or probability model for the data, or using distance measures where objects are a substantial distance from any other cluster. (Han and Kamber 2001)

Outliers that deviate significantly from the group, i.e. with indicator values that varied drastically from returns from other companies with similar profiles, were discarded. For instance, one company had a profile of 2 employees making a 5 million pounds annual turnover. While this may not be impossible, it is far from the general trend. Datasets with same or similar inputs but different outputs, so as this, would confuse the neural network during its learning, so these outliers were removed manually.

- Normalisation

Neural networks, in common with other distance-based mining algorithms, are sensitive to the range of the inputs. For example one input value, the number of computers, spans from 1 to 9000 units (sets), while another input value, the annual turnover value spans from 0 to 1.7 million units (pounds). If these two inputs were fed into the neural network without being normalised, the "annual turnover" input would outweigh the "computer number" input. Thus incorrect weights would be given. The neural network would be misled even if repetitive training might be able to find the right weight in the end. Min-max normalisation, mapping a value $v$ of A to $v'$ was computed using:

$$v' = \frac{v - \min_A}{\max_A - \min_A}(new\_\max_A - new\_\min_A) + new\_\min_A \tag{3}$$

$\min_A$ and $\max_A$: minimum and maximum values of an attribute A

$new\_min_A$ and $new\_max_A$: minimum and maximum values of the new normalised range

In this study, normalisation was used to rescale all the attributes to the range 0 to 1, so the simplified function is:

$$v' = \frac{v - \min_A}{\max_A - \min_A} \tag{4}$$

The min-max normalisation performed a linear transformation on the original data, so as to preserve the relationships among the original data values (Han and Kamber 2001).

### 4.4.2 Missing Values

Missing values was one of the biggest problems in the survey data and one of the most significant challenges to resolve.

The most common approaches for dealing with missing value problem include (Holt and Benfer 2000):

i) Listwise deletions, i.e., the deletion of any case that had one or more missing value. Such deletions would result in a) a smaller sub-sample, that might no longer be representative and b) in variances and covariances that might be considerably reduced.
ii) Pairwise deletions, in which the correlation matrix was computed with all complete pairs of variables. This would help alleviate the problem of sample size reduction. However, since the correlation matrix was generated from different subpopulations with different sample sizes, correlation coefficients and other statistics could be biased.
iii) Insert the mean in place of the missing value.
iv) Use least square multiple regression of available specimens to predict missing values in those cases with incomplete sets of measurements.

Han and Kamber (Han and Kamber 2001) presented similar methods as: ignore the tuple; fill in the missing value manually; use a global constant to fill in the missing value; use the attribute mean to fill in the missing value; use the attribute mean for all samples belonging to the same class as the given tuple; and use the most probable value to fill in the missing value.

Other techniques for dealing with missing value problem adopted by other studies include neural network, iterative regression or hybrid approach (Zhong et al. 2004), (Junninen et al. 2004), (Brouwer and Pedrycz 2003), (Markeya et al. In Press), (Pesonen et al. 1998). It seems that handling missing values has become an independent research field called "data imputation", which has aroused special interest from many researchers.

In this study, it was initially decided to try to develop the neural network, using the complete data sets to train the network, using missing attributes as the outputs and other attributes as the inputs, thus using the neural network to predict the missing values of the incomplete datasets. However, the sample size of the complete datasets was insufficient to train a neural network to forecast. The plan was therefore abandoned.

The primary objective of this study isn't the creation of perfect sets of data. It is to demonstrate how to analyse the relationship between the indicators once data is available. A customised approach was therefore developed in this case. It is acknowledged that this might not be perfect or accurate in the data imputation field. However the approach was judged to be the most suitable one for this study in terms of time consumed to solve the problem and maximum preservation of the data features. The following methods were used to fill the missing values:

i) Using the nearest neighbour – if a company didn't provide a value for electricity consumption and there weren't many companies in the same range, then the missing value was borrowed from the nearest neighbour.

ii) Using the mean of the group – using same example of the above, if there were a few more companies with a similar profile, then the missing value was filled using the mean of the sub-group, assuming the incomplete data sets carried the trends and features of the complete data sets.

iii) Using a correlated value - a correlation matrix of all attributes was calculated based on all complete datasets. Following Han and Kamber (Han and Kamber 2001), the correlation coefficient between attributes A and B was measured by

$$r_{A,B} = \frac{\sum_{i=1}^{N}(a_i - \overline{A})(b_i - \overline{B})}{N\sigma_A\sigma_B} \tag{5}$$

$N$: the number of tuples

$a_i$ *and* $b_i$ : the respective values of A and B in tupple $i$

$\overline{A}$ and $\overline{B}$ : the respective mean values of $A$ and $B$

$\sigma_A$ and $\sigma_B$ : the respective standard deviations of $A$ and $B$

The most correlated attribute was given more weight in considering the missing value. For example, if "employee number" had a higher correlation value to the "electricity consumption" than "computer number", "employee number" was given more importance into filling the missing value of "electricity consumption" rather than "computer number".

Filling the missing values was conducted manually. This was time consuming but it was believed to be the only possible method in this case, considering the quality of the survey data. Also integrating different approaches would avoid the disadvantages of one single approach. For example, if missing values had been replaced by "Unknown" then the mining program might mistakenly have thought that they formed an interesting concept, since they would all have a value in common – that of "Unknown" (Han and Kamber 2001). This kind of hybrid approach (Han and Kamber 2001) used the most information from the present data to predict missing values. The approach was thus expected to have a greater chance to preserve the relationships between the attributes.

However, filling missing attributes with any value inevitably changes the original data to a certain degree. The ultimate challenge was to minimise the effects and preserve the relationships between the variables.

Considering awareness will be raised and more data will be available in the future, this step should not be needed indefinitely.

### 4.4.3 Data Reduction

Data reduction techniques can be applied to obtain a reduced representation of the data set that is much smaller in volume, yet closely maintains the integrity of the original data. Mining on the reduced data set should be more efficient yet produce the same or almost the same analytical results (Han and Kamber 2001). Various strategies for data reduction have different purposes, e.g. data compression is used to reduce the data size. The main aim of data reduction in this study was not to reduce the dataset size since the database wasn't large. Rather, the objective was to reduce the dimension of the datasets so minimum attributes would be required to represent the data features. The size of neural network could thus be minimised so the limited data collected could satisfy the training. The dimension reduction was implemented through fuzzy logic which will be explained in the following section in details.

Also, some factors in a similar category were combined in a point system for simpler cases. For example, whether e-business adoption had increased the supplier choice and customer number, and reduced employee number, supply chain cycle and business transaction cost were joined together to create an e-business point factor. A plus point was given to each answer that represented e-business making a positive contribution. The sum of all these five questions was used as one attribute value for an "e-business points" factor.

Data integration, the merging of the data from multiple data sources, was not appropriate in this study as the two questionnaires were focused on two different industries.

## *4.5 Implementation of Fuzzy Logic*

As stated previously, a neural network only takes numerical data, so fuzzy logic is used in this study to convert the linguistic attributes to numbers a neural network can recognise. On the other hand, it also helps to reduce the number of attributes and training cycles of the network. This section will present how this was implemented.

### 4.5.1 Introduction

To implement the Fuzzy Logic methodology attention was mainly focussed on the data from section 2.7 in the second questionnaire. Twelve aspects were considered against five degrees of change categorised as follows: decreased a lot; decreased slightly; not much change; increased slightly; increased a lot. The aspects considered were, paper consumption, electricity consumption, gas, warehouse, business travel frequency, packaging and so on.

Having these twelve attributes individually presented to the neural network would significantly increase the need for training data, complication of the structure as well as processing speed. Fuzzy logic's role here was twofold: i) convert linguistic descriptions from the questionnaires to numerical data that neural network can read and process, and ii) combine certain attributes to categories so fewer inputs for the neural net and correspondingly smaller size of the network, faster processing and less data required too.

The twelve attributes were categorised according to their effects or relations e.g. electricity, gas and business travel were categorised into one group since they all make contributions to $CO_2$ emission; warehouse space and office space were combined to make land-use output; paper usage and packaging were joint together. In this way seven attributes were reduced to three. Regarding the rest of the five, "employee number" had an absolute value to be input to the neural network. The other four had to be stand-alone as there wasn't any obvious link among them.

Standard fuzzy computing with five steps was applied in the Matlab environment: i) fuzzify inputs, ii) apply fuzzy operators, iii) apply implication method (IF-THEN rule), iv) aggregate all outputs, fuzzy sets that represent the outputs of each rule are combined into one single fuzzy set, v) defuzzify, resolve the fuzzy set, which is a range of output values, to one single output value. (MathWorks 1995-2005)

The five primary Graphic User Interface (GUI) tools for building, editing, and observing fuzzy inference systems in the Fuzzy Logic Toolbox were: the Fuzzy Inference System or FIS Editor, the Membership Function Editor, the Rule Editor, the

Rule Viewer, and the Surface Viewer. These GUIs were dynamically linked, so changes made to the FIS using one of them, would affect any of the other open GUIs.

### 4.5.2 Construction of the Fuzzy Systems

The procedure to implement construction of the fuzzy systems is as follow:

i) Input 'fuzzy' in the Matlab command window, the 'FIS Editor: Untitled' would appear in the new window (Figure 4-11)

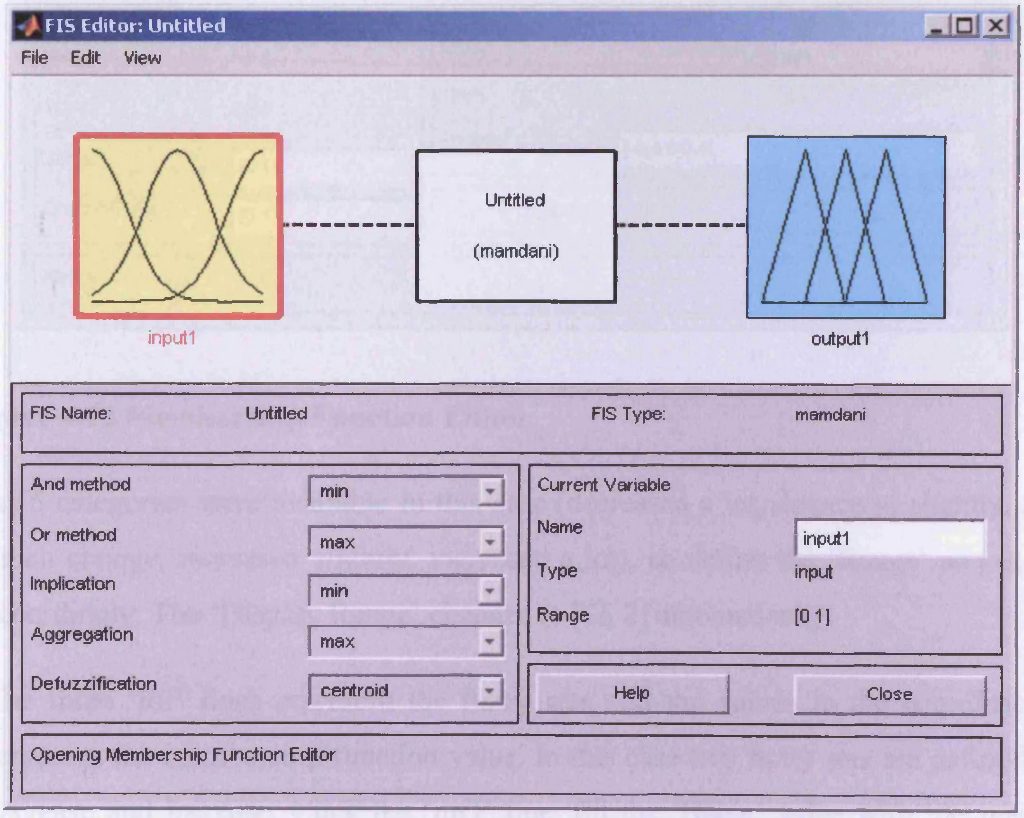



**Figure 4-11 FIS Editor**

ii) Double click 'input1' or choose 'edit'- 'membership functions', the 'Membership Function Editor: Untitled' would appear in a new window (Figure 4-12)

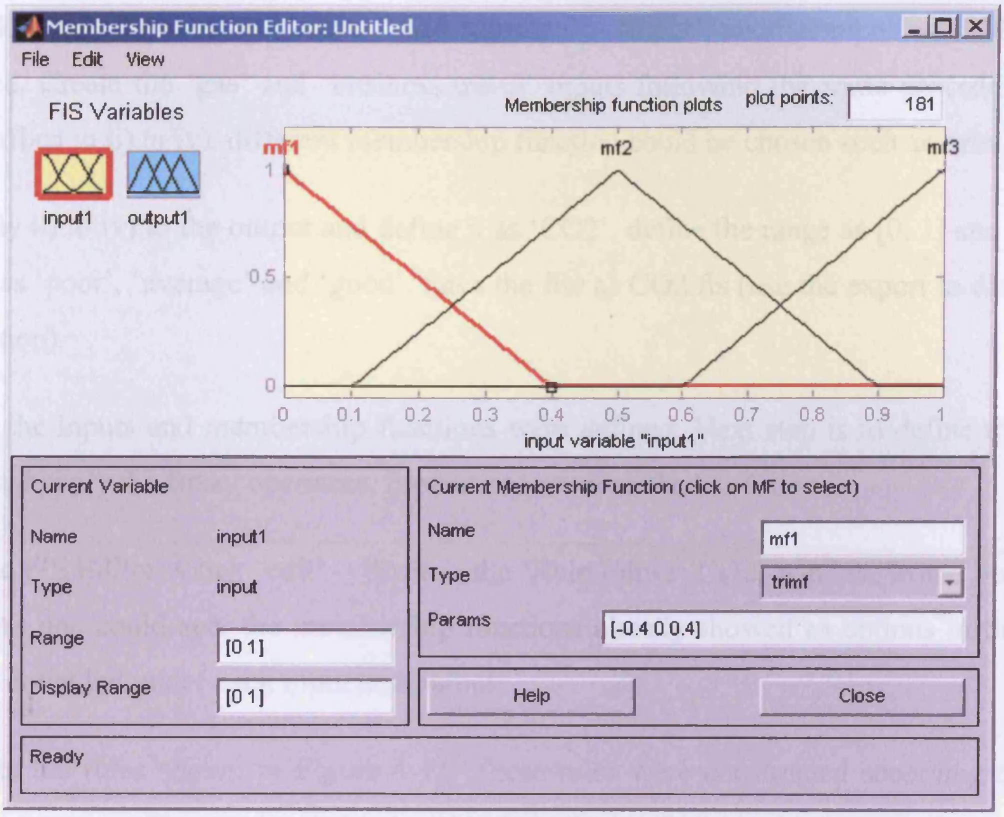



**Figure 4-12 Membership Function Editor**

iii) As 5 categories were available in this case (decreased a lot, decreased slightly, not much change, increased slightly, increased a lot), so define the 'Range' as [-2, 2] accordingly. The 'Display Range' changes to [-2, 2] automatically

iv) The three 'mf' lines represent the fuzzy sets and the values in the coordination represent the membership function value. In this case two fuzzy sets are defined as decrease and increase. Click the 'mf1' line, fill the 'Name' value with 'decrease', and choose type as 'gauss2mf'. Same applies to 'mf3'. The middle 'mf2' line is not in use since we only have two sets so click 'edit'- 'remove selected mf' to remove it. The 'params' is to define the turning point, which is set to [0.2 -2 0.55 -1.8] but can be changed to other sensible values. For example, while [-2, 2] represent [decreased a lot, increased a lot], a value of 0.55 means the company hardly had any change in electricity consumption, thus its membership value to the 'decrease' set (the vertical axis) is approximating 0. Close the Membership Function Editor window, so back to the FIS Editor, change the name of 'input1' to 'electricity'.

v) Click 'edit' in the menu, choose 'Add Variable' – 'Input', another input would be added. Create the 'gas' and 'business travel' inputs following the same procedure described in ii) to iv), different membership function could be chosen such as 'trimf'

vi) Apply ii) to iv) to the output and define it as 'CO2', define the range as [0, 1] and 3 sets as 'poor', 'average' and 'good'. Save the file as CO2.fis (use the export to disk function).

So far the inputs and membership functions were defined. Next step is to define the rules and apply the fuzzy operators. Procedures are described as follow:

vii) In the FIS Editor, Click 'edit' – 'Rules', the 'Rule Editor: CO2' window would pop up. As one could see, the membership functions already showed as options in the drop down list under each input and output

viii) Fill in the rules shown in Figure 4-13. These rules were constructed according to common sense. The number behind each rule in the bracket was the weight value for that rule in the overall rules system. For example, if the three inputs - electricity and gas consumption, and business travel all decreased then performance of $CO_2$ emission was assumed to be good. The weight of that rule was assumed to be 1, which was greater than the 0.75 weight given to Rule 3 where only electricity and gas consumption decreased. These numbers were assumptions and assigned for representing relative significance of each other (that one rule is more significant than the other) and demonstration of the overall methodology. Once more knowledge may be available in the future, these parameter values can always be changed for more accuracy.

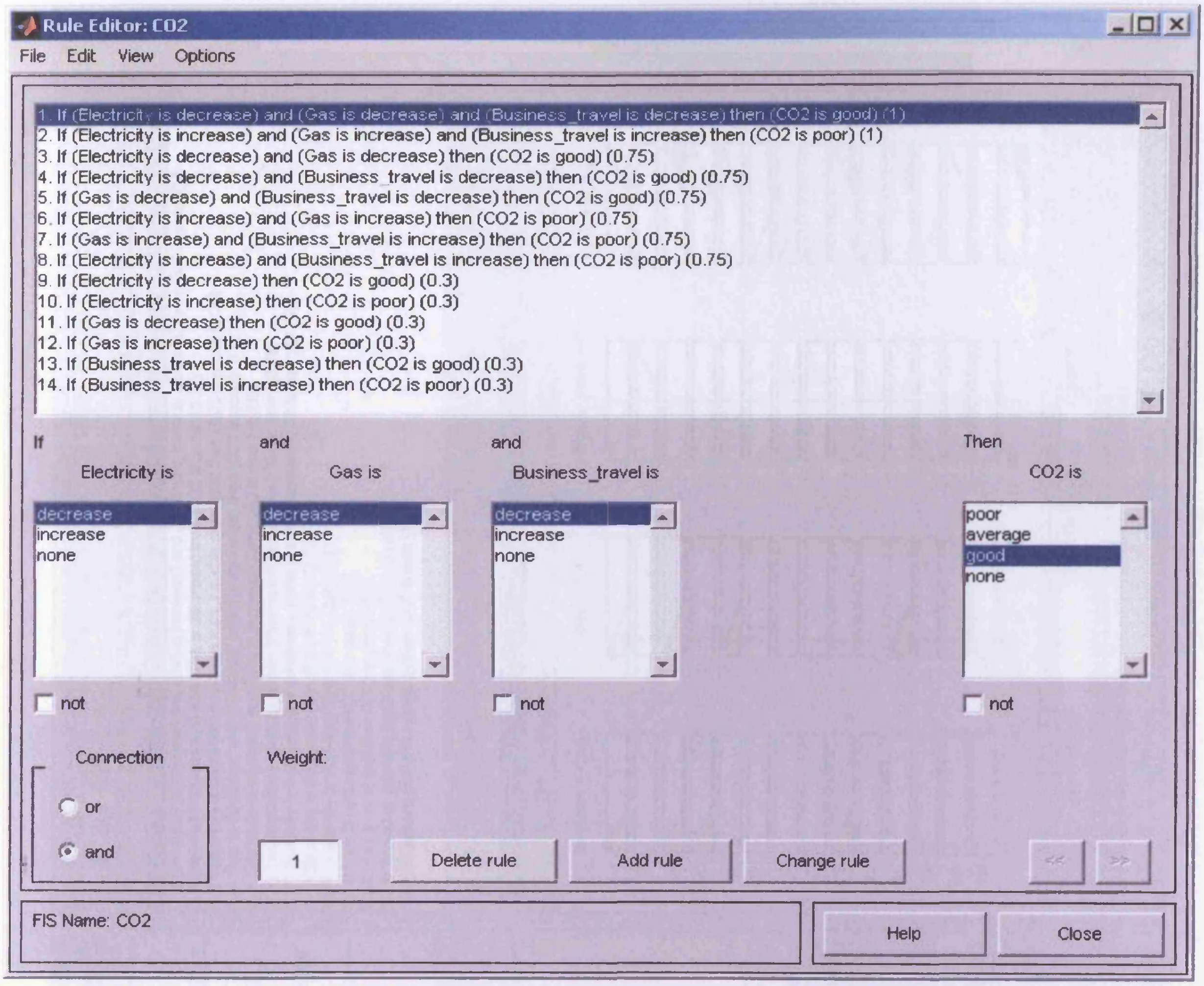



**Figure 4-13 Rule Editor: CO2**

Once this was done, the rules and accumulation result of all the rules could be viewed in the rule view window. Therefore a $CO_2$ fuzzy inference system was built (Figure 4-14).

In the Rule View window, for example, input the 3 inputs values as [-1.64, -1.4, 0], the 14 rules were applied automatically thus a aggregated output value 0.842 was calculated, with [0, 1] representing $CO_2$ emission from poor to good, one could approximately understand the performance level in this case. This numerical number, 0.842, would be the one to represent the three inputs (electricity, gas and business travel) in the neural network system later on.

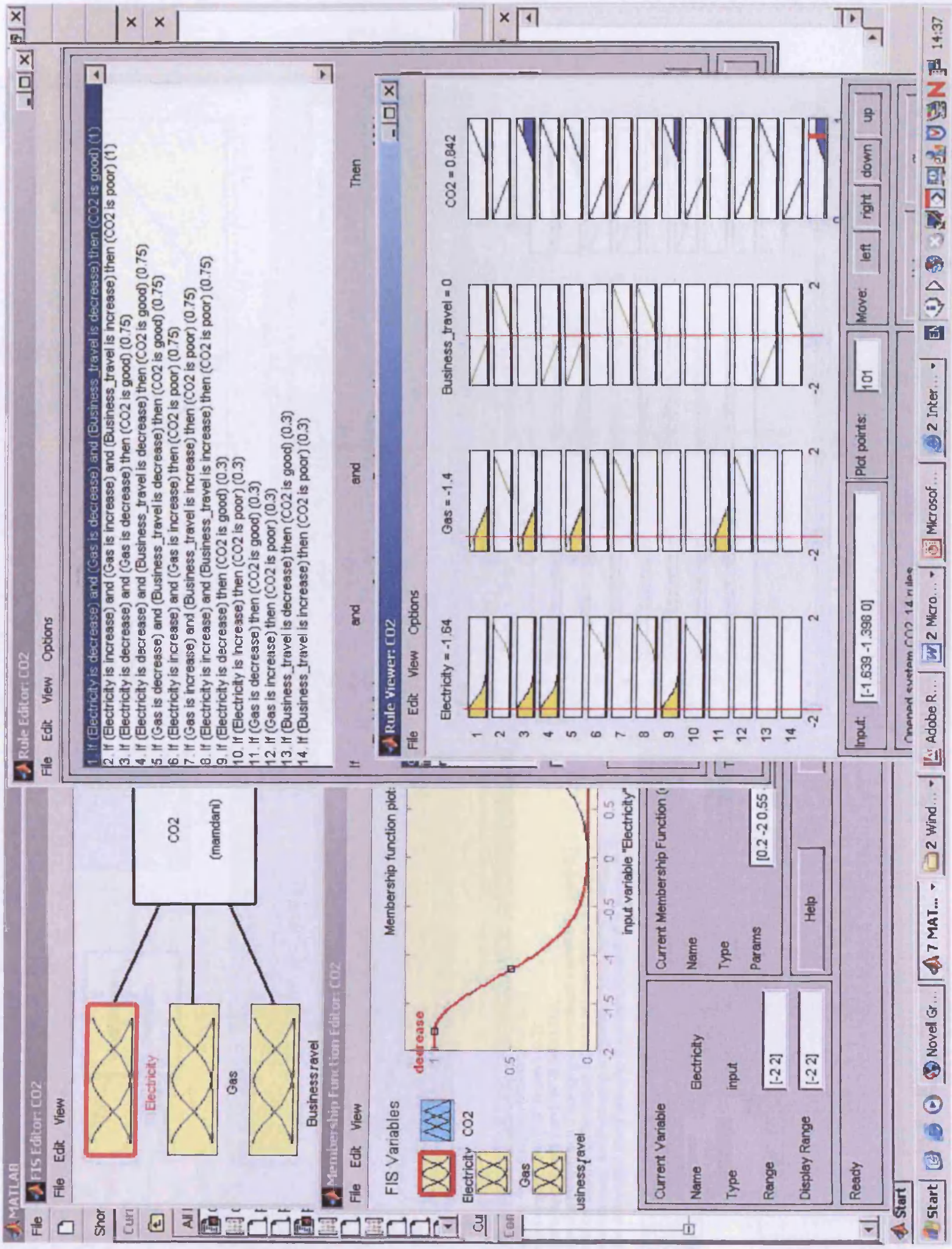


**Figure 4-14 Fuzzy system for $CO_2$**

Following the same procedure, a paper-packaging.fis (Figure 4-15) and landuse.fis (Figure 4-16) were built too.

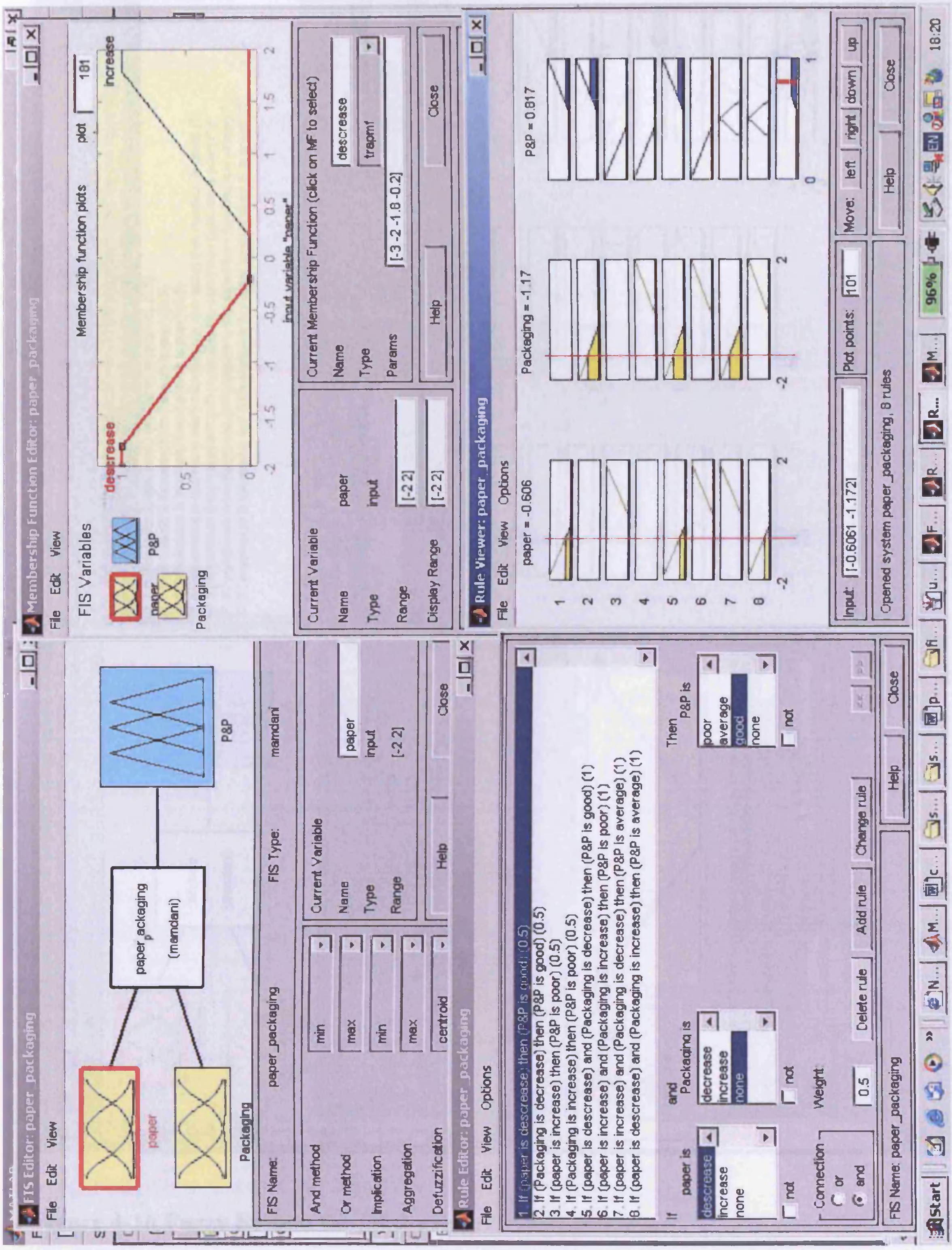



Figure 4-15 Fuzzy System for paper-packaging

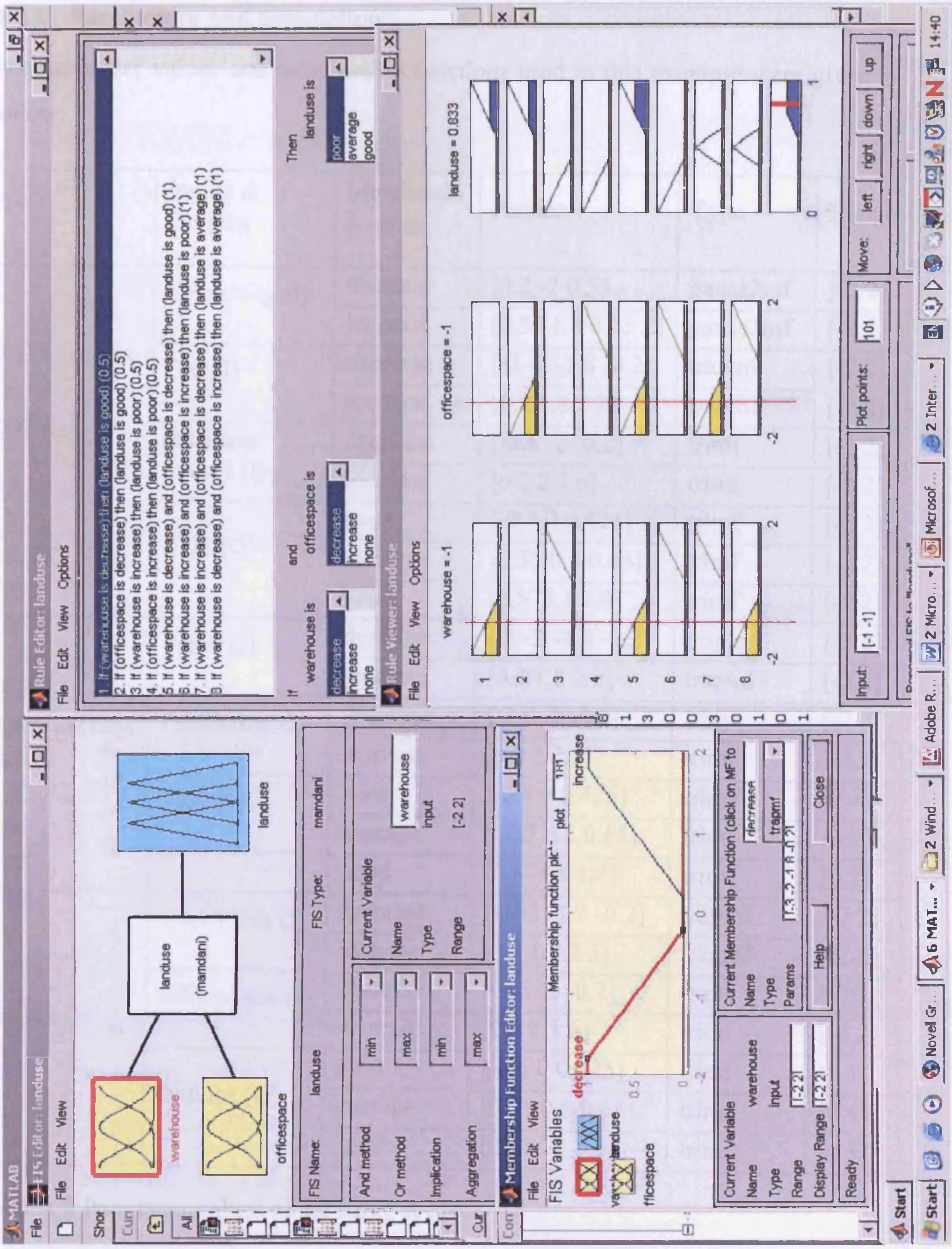


**Figure 4-16 Fuzzy System for land-use**

### 4.5.3 Parameters and Simulation

The parameter values and membership functions used in this example were given as follow:

| FIS | Inputs & Outputs | Membership name | Params | Type | Range |
|---|---|---|---|---|---|
| CO2.fis | Electricity (I) | decrease | [0.2 -2 0.55 - | gauss2mf | [-2 2] |
| | | increase | [0.55 1.8 0.55 2] | gauss2mf | [-2 2] |
| | Gas (I) | decrease | [-2 -2 -1.8 -0.2] | trapmf | [-2 2] |
| | | increase | [0.2 1.8 2 2] | trapmf | [-2 2] |
| | Business Travel (I) | decrease | [-3.6 -2 -0.2] | trimf | [-2 2] |
| | | increase | [0.2 2 3.6] | trimf | [-2 2] |
| | CO2 (O) | poor | [-0.4 0 0.425] | trimf | [-2 2] |
| | | average | [0.35 0.5 0.65] | trimf | [-2 2] |
| | | good | [0.575 1 1.4] | trimf | [-2 2] |
| paper_packaging.fis | Paper (I) | decrease | [-3 -2 -1.8 -0.2] | trapmf | [-2 2] |
| | | increase | [0.2 1.8 2 3] | trapmf | [-2 2] |
| | Packaging (I) | decrease | [-3.6 -2 -0.2] | trimf | [-2 2] |
| | | increase | [0.2 2 3.6] | trimf | [-2 2] |
| | P&P (O) | poor | [-0.4 0 0.425] | trimf | [0 1] |
| | | average | [0.35 0.5 0.65] | trimf | [0 1] |
| | | good | [0.575 1 1.4] | trimf | [0 1] |
| landuse.fis | warehouse (I) | decrease | [-3 -2 -1.8 -0.2] | trapmf | [-2 2] |
| | | increase | [0.2 1.8 2 3] | trapmf | [-2 2] |
| | officespace (I) | decrease | [-3.6 -2 -0.2] | trimf | [-2 2] |
| | | increase | [0.2 2 3.6] | trimf | [-2 2] |
| | landuse (O) | poor | [-0.4 0 0.425] | trimf | [0 1] |
| | | average | [0.35 0.5 0.65] | trimf | [0 1] |
| | | good | [0.575 1 1.4] | trimf | [0 1] |

Table 4-1 Parameter values of the fuzzy systems

The final step was to apply the real data in the three fuzzy systems. In order to batch process the data from the survey rather than one by one, the data was exported from the Microsoft Access database to three individual Excel sheets corresponding to the relevant values of the three fuzzy systems. In the Matlab command window, read the data into the fuzzy systems accordingly, process the calculation (with the 'evalfis' function), and export the outputs directly back into the excel sheets (Figure 4-17).

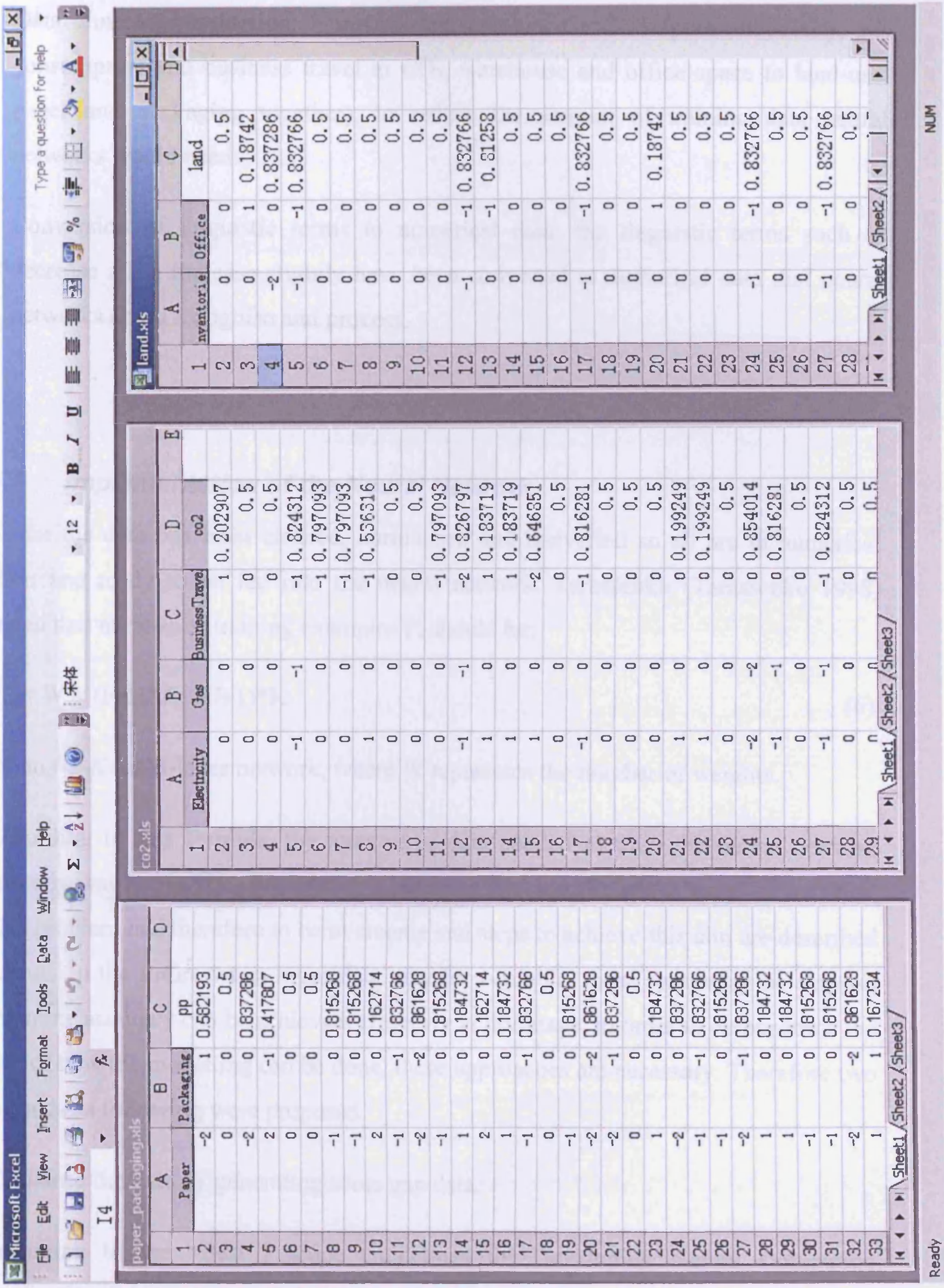



**Figure 4-17 Outputs of the fuzzy systems**

Therefore the implementation of Fuzzy Logic was completed. Two main objectives were achieved:

i) Data dimension reduction: 7 inputs were combined to 3 outputs (electricity, gas consumption and business travel to $CO_2$, warehouse and office space to land-use, paper and packaging together), to reduce the amount of training data neural networks would need.

ii) Conversion of linguistic terms to numerical data: the linguistic terms such as decrease a lot, increase slightly have been converted to numerical data that neural networks could recognise and process.

### 4.6 Implementation of the Neural Network

So far the data has been cleaned, normalised and converted so all are in numerical form and ready to be fed into the neural network. Tarassenko (Tarassenko 1998) stated that number of training examples P, should be:

$$P >= W = (I+1)*J + (J+1)*K \qquad (6)$$

for an *I-J-K* multi-layer network, where W represents the number of weights.

According to this formula, the amount of data needed to train the neural network properly was large. The data available however was insufficient to train the network. This problem had therefore to be overcome and steps to achieve this aim are described below. In the future when more data may be available, these steps may be skipped and more accuracy can be achieved. However at this stage, in order to demonstrate the idea of how the modelling can be done, these approaches are necessary. Therefore two methods as following were proposed.

i) Enhance the data by generating some gap data.

According to the "Best Practice Guidelines for Developing Neural Computing Applications" published by DTI (DTI 1994), the user of a neural network might not have adequate data for training the neural network. But the lack of data could be overcome by applying one's knowledge of the problem to simulate the data. Typical approaches would be massaging existing measured data to cover cases that could not be measured, or constructing full-blown simulation programs to generate training data.

Pursuing this approach, some artificial data were generated. For example, there were 2 companies with greater than 20 and less than 25 employees. A new company with 23 employees was "created" with all the parameter values within the range of the data from these two boundary companies. It was expected that this artificial company still carried the trends and patterns of the group so no additional new information was added to the network.

ii) Using modular neural networks instead of an overall network

As stated previously, the data was insufficient for one large sized network. So the idea here is to break down the one overall network that includes all factors, inputs and outputs, to several modular networks, each dealing with certain specific aspect of the problem. A smaller network would significantly reduce the amount of data needed for the training. However, it's important to choose the right factors to compose the right modular network, otherwise important information might be lost during the process. However, this proposed approach does not follow the traditional neural network application, so it needs to be proven valid before applied. The following section will explain in detail the approach and why it is feasible.

### 4.6.1 Modular Approach

After reducing the data dimension with the Fuzzy Logic approach, there were still more than 20 inputs and outputs, which would make a large neural net with e-business indicators on one side and the environmental indicators on the other. The approach of breaking this large network down, proposed here, was to separate the environmental indicators as outputs of separate networks, and select only inputs that relate significantly more to this certain output, to construct a modular network, using the correlation values. It was expected that main features between these factors would be preserved in these modular networks. Certain information would be lost due to excluding other inputs that might make a small but still considerable contribution to the output. But in this case where the quantity of data was a significant problem for the overall study, some compromise had to be made.

However, the theory needs to be tested first of all. Also defining an appropriate threshold to select certain inputs while ignore the rest was critical too.

### 4.6.2 Testing of the Modular Approach

In order to test whether the modular neural network approach was feasible or not, some testing based on artificial data (Table 4-2) was conducted.

The artificial data was generated manually. 7 inputs were created including: x, 2*x, x+17, $x^2$, y, 3*y and z as shown in column B to H of Table 4-2. Constructed in this way, the first four inputs (x, 2*x, x+17 and $x^2$) were highly correlated to each other, while the last two (3*y and z) were not, as presented in Table 4-3.

The output was calculated using the correlated inputs without interference of uncorrelated y and z.

$$Target = 4*B + 5*C+7*D+9*E = 4*(x) + 5*(2*x) + 7*(x+17) + 9*(x^2)$$

The constants 4, 5, 7 and 9 assigned here were for the purpose of demonstration of this methodology, so other numbers would be equally valid.

Certain noise was added deliberately to mimic the real world data. The noise levels were presented in the last column. 25 sets of data were constructed, 20 of them would be used for training the neural net and 5 of them were separated for the testing later on.

| | A | B | C | D | E | F | G | H | I | J | K |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | ID | x | 2x | x+17 | x2 | y | 3y | z | 4B+5C+7D+9E | ta | (ta ≈ 4B+5C+7D+9E) |
| 2 | 1 | 1 | 2 | 11 | 1 | 22 | 60 | 162 | 100 | 100 | 0.00% |
| 3 | 2 | 5 | 9 | 20 | 29 | 29 | 82 | 178 | 466 | 470 | 0.86% |
| 4 | 3 | 9 | 20 | 26 | 100 | 43 | 132 | 205 | 1218 | 1220 | 0.16% |
| 5 | 4 | 14 | 27 | 32 | 200 | 58 | 178 | 229 | 2215 | 2220 | 0.23% |
| 6 | 5 | 20 | 38 | 40 | 400 | 2 | 8 | 239 | 4150 | 4100 | -1.20% |
| 7 | 6 | 6 | 11 | 22 | 40 | 37 | 120 | 198 | 593 | 600 | 1.18% |
| 8 | 7 | 23 | 49 | 42 | 500 | 5 | 20 | 56 | 5131 | 5100 | -0.60% |
| 9 | 8 | 33 | 65 | 49 | 1000 | 9 | 29 | 78 | 9800 | 9900 | 1.02% |
| 10 | 9 | 45 | 87 | 65 | 2000 | 11 | 38 | 89 | 19070 | 20000 | 4.88% |
| 11 | 10 | 57 | 110 | 78 | 3300 | 18 | 60 | 92 | 31024 | 31000 | -0.08% |
| 12 | 11 | 76 | 149 | 90 | 5776 | 109 | 330 | 108 | 53663 | 55000 | 2.49% |
| 13 | 12 | 82 | 167 | 100 | 6800 | 158 | 480 | 129 | 63063 | 63000 | -0.10% |
| 14 | 13 | 90 | 176 | 119 | 8000 | 184 | 550 | 159 | 74073 | 74000 | -0.10% |
| 15 | 14 | 96 | 195 | 117 | 9100 | 207 | 620 | 11 | 84078 | 84000 | -0.09% |
| 16 | 15 | 109 | 220 | 127 | 11200 | 103 | 300 | 23 | 103225 | 102000 | -1.19% |
| 17 | 16 | 200 | 319 | 229 | 40000 | 89 | 270 | 76 | 363998 | 360000 | -1.10% |
| 18 | 17 | 147 | 300 | 170 | 21500 | 62 | 190 | 33 | 196778 | 200000 | 1.64% |
| 19 | 18 | 187 | 380 | 209 | 35000 | 76 | 220 | 42 | 319111 | 320000 | 0.28% |
| 20 | 19 | 240 | 497 | 267 | 57600 | 98 | 300 | 65 | 523714 | 520000 | -0.71% |
| 21 | | | | | | | | | | | |
| 22 | t1 | 2 | 5 | 20 | 5 | 27 | 89 | 200 | 218 | | |
| 23 | t2 | 11 | 23 | 30 | 130 | 17 | 56 | 198 | 1539 | | |
| 24 | t3 | 50 | 97 | 78 | 2689 | 3 | 10 | 78 | 25432 | | |
| 25 | t4 | 40 | 92 | 60 | 1600 | 2 | 7 | 24 | 15440 | | |
| 26 | t5 | 220 | 400 | 260 | 45000 | 160 | 500 | 71 | 409700 | | |
| 27 | | | | | | | | | | | |

**Table 4-2 Test data for modular network**

| A | B | C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|---|---|
| | *x* | *2x* | *x+17* | *x2* | *y* | *3y* | *z* | *4B+5C+7D+9E* | *ta* |
| x | 1 | | | | | | | | |
| 2x | 0.990997 | 1 | | | | | | | |
| x+17 | 0.99901 | 0.988292 | 1 | | | | | | |
| x2 | 0.953558 | 0.944236 | 0.951377 | 1 | | | | | |
| y | 0.462474 | 0.46967 | 0.466513 | 0.301167 | 1 | | | | |
| 3y | 0.4622 | 0.469125 | 0.466644 | 0.301624 | 0.999617 | 1 | | | |
| z | -0.63225 | -0.64426 | -0.62247 | -0.48635 | -0.30429 | -0.30038 | 1 | | |
| 4B+5C+7D+9E | 0.954491 | 0.945246 | 0.952319 | 0.999995 | 0.303034 | 0.303485 | -0.48813 | 1 | |
| ta | 0.955505 | 0.946818 | 0.953247 | 0.999941 | 0.303111 | 0.303546 | -0.49045 | 0.999959537 | 1 |

**Table 4-3 Correlation of test data**

As presented in Table 4-3, the target output was highly correlated to four of the inputs while significant less related to the other three. So the aim of this test was to see whether using highly correlated factors to construct modular networks would produce satisfactory prediction or not. The satisfaction level was measured by the difference of neural networks' prediction and the targets.

Now the main tasks were to select certain inputs to compose various networks, and train the networks with 20 datasets respectively, use the trained networks on the 5 unseen datasets, compare the results with the targets and see which combinations would have less error.

In this case where 7 variable were available, there would be ($2^7$-1) combinations whether an input would be selected or not to compose the network. E.g. if select any 2 inputs (out of 7 available, B, C, D, E, F, G and H) for the network, the combination include: BC, BD, BE, BF, BG, BH, CD, CE, CF, CG, CH, DE, DF, DG, DH, EF, EG, EH, FG, FH and GH. The number of combinations is 21 (C(7,2)).

The total number of various combinations can also be calculated as C(7,1)+C(7,2)+C(7,3)+C(7,4)+C(7,5)+(C7,6)+C(7,7) = 127 combinations ( $C(n,k) = \binom{n}{k} = \frac{n \cdot (n-1) \cdots (n-k+1)}{k \cdot (k-1) \cdots 1} = \frac{n!}{k!(n-k)!}$ if $n \geq k \geq 0$ , C is the combination notation). The one case where there's zero input was excluded. However it would be time consuming to run all the simulations.

So the L8 matrix (Condra 1993) was used to represent the full combinations (127 situations). The author also added 3 more random combinations. In the matrix, 1 represented that variable was included in the simulation while 2 represented the variable was excluded.

In each simulation, a simple feedforward neural net was built, the 20 sets of data were used to train the network, and 5 sets of unseen data were used to test the trained network. The error of the network's performance on unseen data of each trial was presented in the last column of Table 4-4.

As one can see (Table 4-4), the least errors appeared to be in the trials (trial 2, 10 and 11) where the correlated factors were used to train the network, while the most error appeared in the case (trial 4) where all uncorrelated factors were selected.

| trial | x | 2x | x+17 | x2 | y | 3y | z | ta |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 30.31% |
| 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2.18% |
| 3 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 12.76% |
| 4 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 290.21% |
| 5 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 7.69% |
| 6 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 63.92% |
| 7 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 14.23% |
| 8 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 16.08% |
| 9 | | 1 | 1 | 1 | 1 | | | 14.11% |
| 10 | | 1 | 1 | 1 | | | | 3.73% |
| 11 | 1 | 1 | 1 | 1 | | | | 1.27% |

**Table 4-4 11 Trials on the test data**

Therefore using the correlated factors to construct modular networks instead of an overall one was feasible and sometimes could even provide better performance.

### 4.6.3 Structure of the Three Modular Networks

This section will explain how the modular neural network theory was applied in this study.

Table 4-5 shows the correlation value between the e-business and environmental indicators. Three modular networks were chosen to represent the cases and demonstrate the idea. An electricity module, road travel module and air travel module were selected. $CO_2$ emission from each would be calculated accordingly as another indicator.

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | *EmployeeNurOnlineTran* | | *Region* | *Teleconf* | *workerNun* | *em-tel* | *nputerNum* | *ebPoints* | *Office* | *Electricity* | *long* | *medium* | *short* | *RoadTrave* | *AirTravel* |
| 2 | Employeel | 1 | | | | | | | | | | | | | | |
| 3 | OnlineTrar | 0.16983 | 1 | | | | | | | | | | | | | |
| 4 | Region | 0.19953 | 0.07347 | 1 | | | | | | | | | | | | |
| 5 | Teleconf | 0.30318 | 0.41707 | 0.27321 | 1 | | | | | | | | | | | |
| 6 | Teleworke | 0.53911 | 0.53235 | 0.25228 | 0.69136 | 1 | | | | | | | | | | |
| 7 | em-tel | 0.95879 | 0.01413 | 0.14249 | 0.11259 | 0.27758 | 1 | | | | | | | | | |
| 8 | Computerl | 0.84603 | 0.28269 | 0.2889 | 0.27563 | 0.5178 | 0.79035 | 1 | | | | | | | | |
| 9 | ebPoints | 0.02466 | 0.56852 | 0.11126 | 0.35967 | 0.3851 | -0.10178 | 0.23056 | 1 | | | | | | | |
| 10 | Office | 0.84747 | 0.00129 | 0.10963 | 0.12214 | 0.21328 | 0.89472 | 0.66315 | -0.1243 | 1 | | | | | | |
| 11 | Electricity | 0.94621 | 0.11891 | 0.20713 | 0.21796 | 0.38113 | 0.95072 | 0.89005 | 0.03264 | 0.89576 | 1 | | | | | |
| 12 | long | 0.31717 | -0.03609 | 0.59911 | 0.17785 | 0.16285 | 0.30684 | 0.33855 | -0.07448 | 0.23373 | 0.31192 | 1 | | | | |
| 13 | medium | 0.4455 | -0.01119 | 0.55667 | 0.02007 | 0.21074 | 0.43707 | 0.43152 | -0.03749 | 0.41339 | 0.43312 | 0.36031 | 1 | | | |
| 14 | short | 0.76127 | -0.12691 | -0.00647 | -0.0596 | 0.0974 | 0.83548 | 0.51979 | -0.24487 | 0.7311 | 0.73345 | 0.18311 | 0.20045 | 1 | | |
| 15 | RoadTrave | 0.83807 | -0.07444 | 0.08306 | -0.0203 | 0.15408 | 0.90397 | 0.66471 | -0.21827 | 0.89739 | 0.86213 | 0.31899 | 0.40759 | 0.79637 | 1 | |
| 16 | AirTravel | 0.25128 | -0.29878 | -0.15617 | -0.15561 | -0.1806 | 0.34754 | 0.071 | -0.33713 | 0.33712 | 0.23542 | -0.05204 | 0.0627 | 0.36829 | 0.4207 | 1 |

**Table 4-5 Correlation of the survey data**

For the electricity module, the most correlated factors chosen were "employee number", "employee number - teleworker number" (as there were many zeros in "teleworker number"), "computer number" and "office space".

In the air travel module, the threshold for including and excluding factors was not obvious, so the most sensible factors were chosen ("region", "teleconferencing" and "employee number" were more likely to be the influencing factors). For the "teleconferencing hour" and "air travel distance", the variance was rather large and the consistency of the data was not particularly good. They were therefore converted to "teleconferencing scale" (Table 4-6) and broken down to frequencies of long-haul, medium-haul and short-haul travels.

| Teleconferencing hours | Scale |
|---|---|
| 0 | 0 |
| (0,50] | 1 |
| (50, 100] | 2 |
| (100, 250] | 3 |
| (250, 500] | 4 |
| (500, ∞) | 5 |

**Table 4-6 Teleconferencing Scale**

At this stage, the author believes these enhancements of data were necessary for the progress of the overall study.

It was therefore decided that the overall network would be broken down to three modular networks and the structures were decided. Training data requirement was significantly and effectively reduced due to the much smaller size of the networks. Furthermore, data was separated to two parts: one for the training process, and the other retained unseen to the network, for the testing and validation phase.

### 4.6.4 Construction, Training and Testing of the Neural Networks

So far this section has introduced the modular network approach proposed by the author, using correlation coefficient values; the approach was tested to be effective in general; and the structures of the three modular networks were also defined for this study. This following subsection will explain in detail the final implementation of this approach, including construction, training and testing of three modular networks.

The procedure adopted for the building of the three modular neural networks was as follow:

i) First of all, create a folder containing all the data, data exported from MS Access database to MS Excel sheets. Add the path of this folder in the Matlab directories and make it the current directory to work on.

ii) Read the data in Excel to the Matlab workspace in batch (with command 'xlsread'). It needs to be noted that the way Matlab reads a matrix was sometimes different to the Excel structure, so one might need to transpose the matrix. For example, when the minmax normalisation was applied, it should have been applied to one variable, i.e. employee number, from N companies/datasets. However from the way Matlab imported data from Excel, if no transpose was applied, the minmax function would read the minimum and maximum value of 4 variables from one company, which was wrong.

iii) As it was unknown which type of network would provide the best performance, create different networks, including radial basis network (newrb), feedforward (newff), backpropagation.

iv) Define different network structures, which means different amount of hidden neurons.

v) Define the transfer functions, such as logsig and tansig.

vi) Define various parameters, mainly including training epochs, goal and learning rate, training function (train or adapt).

vii) Present the data to the defined network and train the network with the defined training parameters. The training would stop once the performance goal was met or maximum of epoch reached.

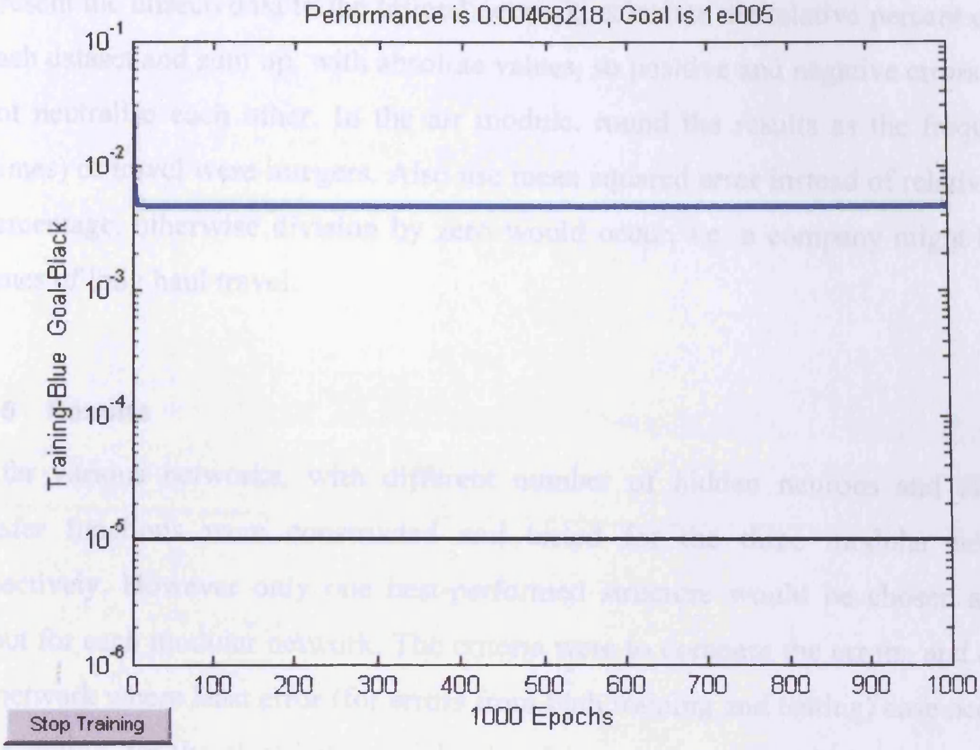



**Figure 4-18 Electricity module training performance – 1000 epochs**

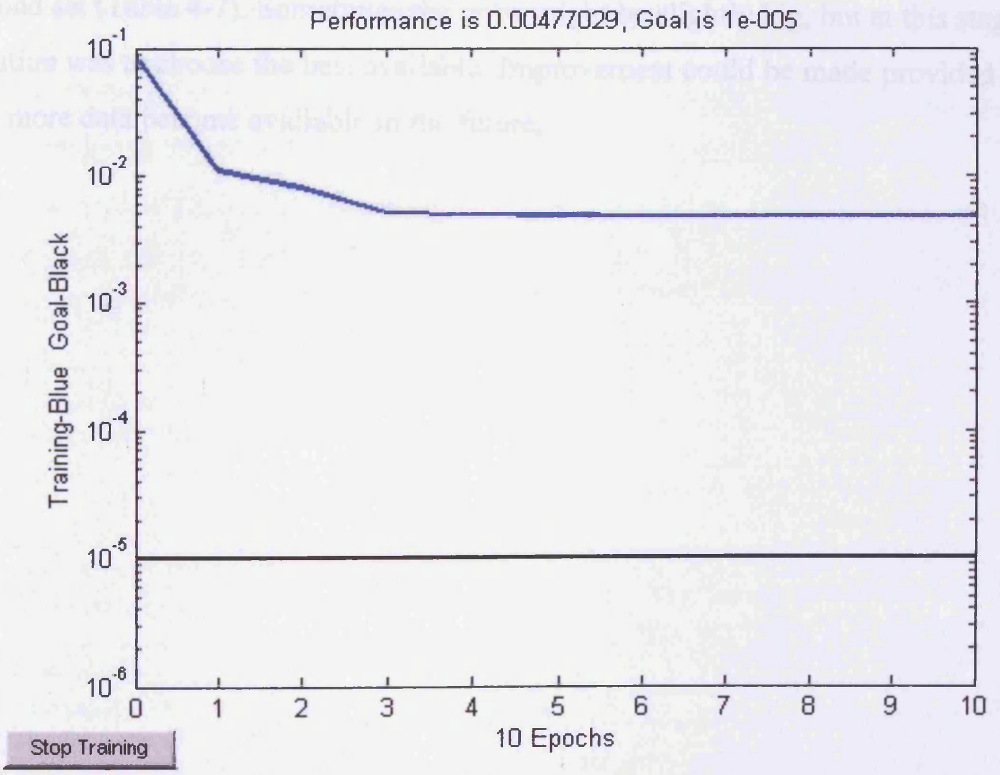



**Figure 4-19 Electricity module training performance – 10 epochs**

viii) Present the unseen data to the trained network, calculate the relative percent error of each dataset and sum up, with absolute values, so positive and negative errors would not neutralise each other. In the air module, round the results as the frequencies (times) of travel were integers. Also use mean squared error instead of relative error percentage, otherwise division by zero would occur, i.e. a company might have 0 times of long haul travel.

### 4.6.5 Results

So far various networks, with different number of hidden neurons and different transfer functions were constructed and tested for the three modular networks respectively. However only one best-performed structure would be chosen as final output for each modular network. The criteria were to compare the errors, and choose the network where least error (for errors from both training and testing) case occurred. For example, for the electricity module, the chosen structure would be 4 (inputs) – 3 (hidden neurons, transfer function: tansig) – 1 (transfer function: purelin) – 1 (output), where the training error was 8.8%, error for first testing set was 4.86% and 3.28% for second set (Table 4-7). Sometimes the error might be slightly big, but at this stage, the solution was to choose the best available. Improvement could be made provided better and more data become available in the future.

| | Targets | Predictions | Errors |
|---|---|---|---|
| Test Set1 | 5.63 | 5.95 | 5.61% |
| | 13.67 | 14.63 | 7.05% |
| | 12.13 | 12.35 | 1.80% |
| | 21.07 | 21.40 | 1.57% |
| | 39.75 | 38.87 | -2.21% |
| | 43.83 | 43.07 | -1.75% |
| | 50.13 | 48.50 | -3.26% |
| | 55.07 | 51.59 | -6.31% |
| | 68.67 | 78.95 | 14.98% |
| | 65.00 | 62.39 | -4.01% |
| | Average | | 4.86% |
| | | | |
| Test Set 2 | 6.24 | 6.30 | 0.86% |
| | 12.08 | 12.64 | 4.61% |
| | 15.95 | 16.61 | 4.13% |
| | 17.70 | 18.45 | 4.26% |
| | 40.17 | 39.09 | -2.68% |
| | 32.37 | 31.59 | -2.41% |
| | 51.67 | 51.70 | 0.06% |
| | 54.67 | 52.02 | -4.85% |
| | 66.67 | 68.65 | 2.98% |
| | 71.00 | 75.25 | 5.98% |
| | Average | | 3.28% |

**Table 4-7 Electricity module results**

Finally export all the parameter values in the trained neural net so a static network could be constructed, as it was found that resetting these initial values in the same workspace in Matlab sometimes didn't really refresh the structure so different or accumulated results might occur. The architecture of the road travel module, for example, was:

*Network structure*
*Road 20 epochs*
*net_road =*

*Neural Network object:*

*architecture:*

*numInputs: 1*
*numLayers: 2*
*biasConnect: [1; 1]*
*inputConnect: [1; 0]*
*layerConnect: [0 0; 1 0]*
*outputConnect: [0 1]*
*targetConnect: [0 1]*

```
    numOutputs: 1  (read-only)
    numTargets: 1  (read-only)
  numInputDelays: 0  (read-only)
  numLayerDelays: 0  (read-only)

  subobject structures:

       inputs: {1x1 cell} of inputs
       layers: {2x1 cell} of layers
      outputs: {1x2 cell} containing 1 output
      targets: {1x2 cell} containing 1 target
       biases: {2x1 cell} containing 2 biases
   inputWeights: {2x1 cell} containing 1 input weight
   layerWeights: {2x2 cell} containing 1 layer weight

  functions:

      adaptFcn: 'trains'
       initFcn: 'initlay'
    performFcn: 'mse'
      trainFcn: 'trainlm'

  parameters:

    adaptParam: .passes
     initParam: (none)
   performParam: (none)
    trainParam: .epochs, .goal, .max_fail, .mem_reduc,
            .min_grad, .mu, .mu_dec, .mu_inc,
            .mu_max, .show, .time

  weight and bias values:

          IW: {2x1 cell} containing 1 input weight matrix
          LW: {2x2 cell} containing 1 layer weight matrix
           b: {2x1 cell} containing 2 bias vectors

  other:

      userdata: (user stuff)

=====================================
net_road.IW{1,1} =
  -0.9363  -1.8916  -1.9167   2.7480
   1.9023  -1.8301  -2.9444  -0.2066
  -3.4389   0.7282   1.7315  -0.5708
  -2.5884  -1.6329  -0.3935   2.4817
net_road.IW{2,1} =    []
-----------------------------
net_road.b{1} =
   3.9598
  -1.3199
```

```
  -1.3199
  -3.9598

net_road.b{2} =   -0.9607
-------------------------------------------
net_road.LW{1,1} = []
net_road.LW{1,2} = []
net_road.LW{2,2} = []
net_road.LW{2,1} = 0.0503  -0.5947   0.3443   0.6762

==================================================
after training
net_road.IW{1,1} =
 -45.5242  81.4700   2.8699 -54.8121
 -45.9917  21.4087  -9.9876 -41.8836
   0.3607  -1.2531   0.2346  -0.8943
   2.0671   4.3992  24.8885  33.6262
net_road.IW{2,1} = []
-------------------------------------
net_road.b{1} =
  15.6411
   9.6372
  -1.5016
 -48.6610

net_road.b{2} =   0.0143

----------------------------------------
net_road.LW{1,1} = []
net_road.LW{1,2} = []
net_road.LW{2,2} = []
net_road.LW{2,1} =   0.4838  -0.4001  -2.0761   4.5656

==================================================
minp_road =
    1
    0
    1
    5

maxp_road =
        100
         98
        120
       1200

mint_road = 0
maxt_road =  15000

minmax(pn_road) =
   -1   1
   -1   1
   -1   1
```

*-1 1*

So far, the three optimal modular networks were found, where the structure and parameters were defined, and ready to work on unseen cases and make predictions.

## 4.7 Conclusions

This chapter has attempted to explain in detail how the data presented in the previous chapter was further processed, mined and modelled. This part of the work forms the engine of this study's Expert Decision Support System.

The methodologies adopted in the study include general data mining techniques, fuzzy logic and neural network. Their concepts, major components and principles were introduced. Also the reasons why these approaches were chosen and how these methodologies fit into this study were discussed.

Data mining, including data cleaning, data imputation and data normalisation were applied to pre-process and enhance the survey data collected.

The software choices for the application were narrowed down to three options, including Matlab, Weka and Clementine. After comparison, Matlab (version 7) was chosen to be the main software programme for this part of work. It includes both a fuzzy logic toolbox and a neural network toolbox.

Fuzzy logic was applied to convert linguistic terms to numerical numbers and to reduce the data dimensions, so less data would be needed for neural network training. Detailed implementation procedures using the Matlab Fuzzy Logic toolbox have been given.

Due to insufficiency of data, an innovative modular neural network approach, based on correlation coefficients between factors, was proposed. The reason why this was chosen over one overall neural network that includes all factors was discussed. Having realised it is unconventional, the author tested the approach on the sample data. The results proved the methodology to be feasible and appropriate for this study.

Finally, three modular networks were selected, defined, constructed, trained, tested and validated. Detailed implementation procedures were presented, reasonableness of the results was discussed, and the optimal network structures were found. This part of work will be used in the forward chaining system of the Expert Decision Support System, as explained in the next chapter.

# Chapter 5

# Expert Decision Support System

## 5.1 Introduction

The previous chapter has explained the development and validation of three neural networks. These three neural networks form the engine of the calculation and prediction of environmental performance indicator values based on e-business indicators. However presentation of the neural networks is in the form of computing codes – functions, callbacks etc built in Matlab. Direct operation of this would involve understanding of Matlab commands and programming. Outputs generated are also in various computational data formats that are not easy to understand. The end-users would be expected to understand how the functions, neural nets, and all the background programming work to be able to use it, which is not practical. Therefore a Decision Support System, built around these three neural networks but has a user-friendlier interface and being able to post-process data to information, is needed.

Also to further assist decision-making, information regarding what's in common among companies who have achieved similar environmental performances is important. For example, one group of companies that were categorised together, has only 1 times*persons long haul air travel per year. After analysis, it's found that "teleconferencing hours" - one of the e-business indicators from these companies fall within certain range. This information can help companies decide what boundary conditions (e.g. range of teleconferencing hours) are needed to achieve certain targets (e.g. one time*person long haul air travel per year). This knowledge discovery process is generally done by Expert System technology, such as rule induction. And the knowledge base and inference rules together form a typical expert system. Therefore typical ES techniques were applied in this study for the purpose of defining targets-conditions relationship between environmental indicators and e-business indicators.

As a result this chapter will explain the development and implementation of the Expert Decision Support System (EDSS) – a hybrid system with two-way simulations. The EDSS is the final interactive presentation system for the end-users of this study.

The forward chain of the EDSS is based on Decision Support System (DSS) technology, with the engine built on the implementation of the data mining, Fuzzy Logic and Neural Network explained in the previous chapter. This part of the system deals with cause-effect (cause: e-business indicators; effect – environmental indicators) calculation, estimation and prediction based upon databases and models.

The backward chain of the EDSS is based on Expert System (ES) technology, with the inference engine built on rule induction and knowledge discovery that will be explained in this chapter. This part presents conditions and rules to be met in order to reach certain pre-defined targets (conditions: e-business indicator range; targets: environmental indicator goals).

The following sections will explain:

- The concept of Expert System in general and how the technology fits into this study.
- The concept of rule induction and knowledge engineering, the implementation process, and the implementation software programme.
- Decision Support System technology and how neural networks contribute to a DSS.
- Why a hybrid system was needed in this study, and programming of the ES inference engine and DSS within Matlab.
- Development of the Graphical User Interface.

- The forward and backward chain simulations of the EDSS, and testing of the EDSS.
- Conclusions.

## 5.2 Decision Support System

This section will explain the concept of a decision support system in general, its standard essential components, and how a DSS is developed and customised in this study.

### 5.2.1 The Concept and Components

The concepts involved in Decision Support Systems were first articulated in the early 1970s by Michael S. Scott Morton (Scott Morton 1971) under the term 'management decision systems' (Turban 1988). Due to its relatively short history, there is no universal standard definition or generic model. The concept varies greatly: in the narrow sense, it can be defined as "an interactive, flexible, and adaptable computer-based information system, especially developed for supporting the solution of a non-structured management problem for improved decision making. It utilizes data, provides an easy-to-use interface, and allows for the decision maker's own insights" (Turban 1995). Adopting a broader definition yields "a computer-based system that aids the process of decision making" (Finlay 1994).

In this study, a DSS is defined as an interactive computer system designed to assist decision-making process with the use of certain analytic methodologies or models. In this case it has the modular neural networks as the engine.

According to Mintzberg et al (Mintzberg et al. 1976), decision-making involves three phases. During the first phase, intelligence, the decision maker searches for conditions calling for a decision. The decision maker may react to problems or may recognise opportunities. Design is the second phase of decision-making. During design, the decision maker develops and analyses alternative courses of action by either searching for ready-made alternatives or developing a custom-made solution. The third phase is choice, during which the decision maker selects the best alternative.

The standard essential components of a DSS include (Schultheis and Summer 1998):

- *A data warehouse/database*: the process of creating a data warehouse is straightforward. Firstly, the data are 'scrubbed' to ensure that they are meaningful, consistent and accurate. Then, the data are loaded into relational tables so that they can be used to support both analysis and query applications. Ultimately, the data need to conform to the logical data model that has been established.

- *Data mining and intelligent agents*: once the data warehouse is created, decision makers need to use tools, which are referred to as intelligent agents, to access and query the data they need. This process is called data mining. Intelligent agents are pieces of software that users can use to analyse trends, to identify exception conditions and to track results. Data mining tools can also be used to identify patterns in the data, to infer rules from these patterns and to refine these rules based on the examination of additional data.

- *Model component*: the goal of model management is to help decision makers to understand a phenomenon about which they are making a choice. The challenge of creating a DSS entails knowing what models to include and knowing how to make them meaningful for the decision maker.

### 5.2.2 DSS in this Study

In the case of this study, the proposed DSS consists of a user interface, databases, a data-mining engine and models based on Fuzzy Logic and Neural Networks. To operate the system, the user inputs certain parameters, including general company profile information such as employee number, business region, etc. The system will then ask the user to provide further e-business parameter values, such as how many computers the company has, how often they use teleconferencing, how many employees work from home, etc.

After all the data are provided, the information will be fed into the core and engine of the DSS. The system will then locate the most matching pattern and present the average figures in this industry/sector. As an output, the system will tell the user how environment friendly they are compared with similar companies who applied or limited certain e-business technologies.

Also, the user can provide his/her existing environment-related figures, such as business travel data, annual energy consumption data etc, and compare these with the estimations made by the neural network that represent the average performance in the industry sector. Current strategies could be evaluated and changes could be suggested.

Furthermore, the user can adjust one or more parameter values while keep the rest unchanged, and simulate again. He/she can see how the proposed changes could potentially affect the outcome. For example, filling different values for numbers of teleworkers while keeping other variables constant, see how this would affect the electricity consumption and corresponding costs, or see how business travel and related $CO_2$ might change with different usage of tele-conferencing.

Hereby, in the scope of this study, a DSS could be beneficial to help an individual company to be more environmental friendly, by applying or limiting certain e-business operations. Certain simulations in the DSS can help the company decide how to do what, to balance its environment friendliness and profitability.

Ideally, the DSS should be integrated into the company's exiting systems such as Management Information System (MIS), Environmental Management System (EMS), Budgeting System, etc.

## 5.3 The Hybrid Approach

This section will explain the concept of an Expert System in general, its major components, the process of building an ES, the differences between DSS and an ES, and why a combination of both is needed in this study.

### 5.3.1 Expert System

As presented, the forward chaining DSS is based on Neural Network technology. This forward chaining logic can analyse the cause-effect relationship of e-business/ICT and the environment, but not the reverse conditions-targets relationship. This means in the simulation the inputs and outputs cannot be switched. So in the case one wants to know how to set the e-business indicators values (inputs) to achieve certain environmental performance goals (outputs), the DSS can not really help, unless the user runs the simulation with random inputs in the neural networks till she/he finds the answers, which could be time-consuming. In order to simulate the conditions-targets relationship, for example, in order to achieve 1 times*persons/year long-haul air travel, how many teleconferencing hours should be proposed, an expert system with a knowledge base and inference engine (e.g. If... Then... Rules) is therefore needed.

Expert systems are computerised advisory programs that attempt to imitate or substitute the reasoning processes and knowledge of experts in solving specific types of problems. This possibility may have a significant impact both on advisory professionals (financial analysts, lawyers, tax advisers, etc.) and on organisations and management. ES is a branch of applied Artificial Intelligence and was developed by the AI community in the mid-1960s (Turban 1988).

The major components of an expert system (Figure 5-1) include (Schultheis and Summer 1998):

i) A knowledge base, which contains the information and the rules of thumb that the ES uses to make decision.

ii) The inference engine, the central processing unit of the ES, which conducts the dialogue with the user, asks for information and applies it and uses the knowledge base to draw conclusions for each situation.

iii) The knowledge acquisition subsystem, where new rules can be added to the knowledge base.

iv) The explanation subsystems, which explains the procedures that are being used to reach a decision.

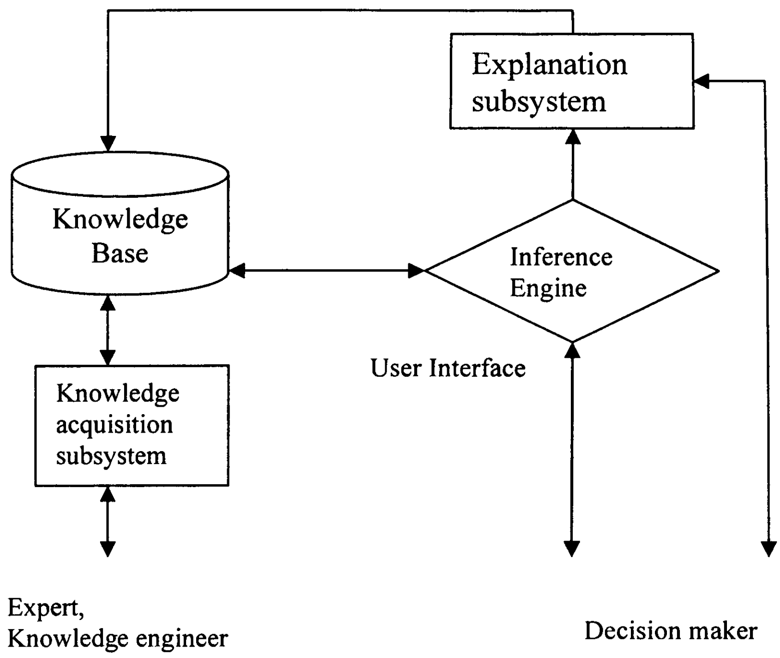



**Figure 5-1 Expert System components (Schultheis and Summer 1998)**

The process of building an expert system is called knowledge engineering. The customary way of building an expert system, constructing a prototype, testing, interviewing, and so on, is a time-consuming and labour-intensive task. The problem of transferring human knowledge into an expert system is so major that it is called the knowledge acquisition bottleneck (Giarratano and Riley 1998).

The typical knowledge engineering process stated by Turban includes: (Turban 1988)

i) Knowledge Acquisition, the first step of knowledge engineering (Feigenbaum and McCorduck 1983), is the extraction of knowledge and involves the acquisition of knowledge from human experts, books, documents, sensors, or computer files.

ii) Knowledge Representation, encoding the knowledge in the knowledge base so that appropriate inferences can be made.

iii) Inference, the design of the software that will enable the computer to make inferences based on knowledge, and then provide advice to the user on specific issues.

iv) Explanation and Justification, the design and programming of an explanation capability.

The major methods of knowledge acquisition are interviews, protocol analysis, observations, and rule induction, while the rule induction is the one chosen for this study. Induction means a process of reasoning from the specific to the general. In ES terminology it refers to the process in which rules are generated by a computer programme from example cases. A rule-induction system is given examples of a problem where the outcome is known. When it has been given several examples the rule-induction system can create rules that are true for the example cases. The rules can then be used to assess other cases, where the outcome is not known (Turban 1988).

### 5.3.2 Expert Systems versus Decision Support Systems

Decision support systems and expert systems are different in many senses, including their components, processing, structures etc. as presented in Table 5-1. In this study, they complement each other, fulfilling different purposes, which will be explained further in the next section.

The forward chaining DSS in this study is more of a data processing system. In a data processing system, the data structures are defined and programmed code is used to express procedures. These procedures are executed in the same way and in the same order on all the data that are processed. In contrast, an expert system does not execute its rules in the same order. When an expert system receives input data, it selects the rules that apply to the problem. As it asks the user additional questions, it learns more about the situation and applies further rules. (Schultheis and Summer 1998)

| | Decision Support Systems | Expert Systems |
|---|---|---|
| Objective | Assists the human | Provides "expert" consulting |
| Decision-maker | The human | The system |
| Query type | Human queries the machine | Machine queries the human |
| Problem area | Complex, wide | Narrow domain |
| Database | Includes factual knowledge | Includes procedures and data |
| Evolution | Adapts to the changing environment | Supports a fixed problem domain |
| Design process | Adaptive design process, involves a user, who identifies an information need, and a designer, who develops an initial version of the proposed system and continues to modify it as the user's needs change. | Involves a knowledge engineer (KE) and an expert. The KE may take systematic interviews over a period of months to acquire knowledge from an expert. The KE must decompile or break down the expert's knowledge into the hundreds of rules that are part of a complex reasoning process. |
| Tools | Database query, modelling, data analysis, display software | Problem-oriented languages such as LISP or PROLOG, these are symbol-manipulation languages |
| Examples | A good example of a DSS would be one that would help in making the decision to take a four wheel drive truck or a two wheel drive truck on a camping trip. | An example of an expert system would be one that would determine if one had a certain disease or not. |
| Structure | A DSS involves a series of logical queries that ultimately leads to the best possible answer | An expert system involves a series of true or false statements that ultimately leads you down a hierarchical tree structure to the correct answer. |
| Users | Interactive computer-based systems, primarily used by top and middle management | Systems that make routine decisions, decisions that are made over and over again, usually utilized by middle and operational management |

**Table 5-1 DSS vs. ES**

### 5.3.3 The Combination

There are two main methods of reasoning when using inference rules: backward chaining and forward chaining (Wikipedia 2006):

Forward chaining starts with the data available and uses the inference rules to conclude more data, until a desired goal is reached. An inference engine using forward chaining searches the inference rules until it finds one in which the if-clause is known to be true. It then concludes the then-clause and adds this information to its data. It would continue to do this until a goal is reached. Because the data available determines which inference rules are used, this method is also called *data driven.*

Backward chaining starts with a list of goals and works backwards to see if there is data which will allow it to conclude any of these goals. An inference engine using backward chaining would search the inference rules until it finds one which has a then-clause that matches a desired goal. If the if-clause of that inference rule is not known to be true, then it is added to the list of goals.

DSS and ES are normally independent systems. However, in this study, they were integrated as an Expert Decision Support System, a hybrid system, to fulfil these two chaining logics.

The hybrid approach is an interactive system with two-fold meanings. On one hand, the forward chaining decision support system, from inputs to outputs, is a simulation. Users can change parameters, and see how this will affect the outputs, which are the environment impact indicators. On the other hand, the backward chaining expert system can help users decide how to change certain parts of their business by giving conditions or ranges of certain input, to achieve the targeted goals. These could include change to be more cost-effective or more environmental friendly.

The two systems were not built following the standard DSS and ES development approach. However, they incorporated the major components and concepts of DSS and ES. The implementation in this study was derived from the original theories of DSS and ES. Certain standard procedure and components were extended for better integration with each other.

## 5.4 Rule Induction and Knowledge Discovery

As stated before, the major methods of knowledge acquisition are interviews, protocol analysis, observations, and rule induction. Induction means a process of reasoning from the specific to the general. In ES terminology it refers to the process in which rules are generated by a computer programme from example cases. A rule-induction system is given examples of a problem where the outcome is known. When it has been given several examples the rule-induction system can create rules that are true for the example cases. The rules can then be used to assess other cases, where the outcome is not known (Turban 1988). For this capability, rule induction is chosen for this study.

In order to build the expert system, a rule inference system and knowledge base are needed. In a standard case, the system development would involve interaction among expert, knowledge engineering and end-users. However since this is a new research field, the knowledge discovery in this case, depended on the rule induction on the survey data.

The rule induction and knowledge acquisition are actually still part of data mining, but different to the techniques used in previous chapter, the focus of this part of data mining was to categorise the data and extract rules and patterns.

### 5.4.1 Clementine and C5.0 Algorithms

The previous chapter has introduced the software programmes including Clementine for the Neural Network part of work. Matlab was chosen over Clementine because of its powerful computing environment and better availability of neural network algorithms and functions. However, for the knowledge engineering, Clementine shows particular strengths in many aspects of knowledge engineering, including various clustering methods such as C5.0 and K-means, rule induction, decision tree etc. Its data mining techniques are straightforward and user-friendly.

In Clementine, the rule induction algorithm is called C5.0, a commercial Windows95/NT decision tree and rule induction product from RuleQuest (Quinlan 2000b) developed by Ross Quinlan as the successor to his successful and widely used

ID3 and C4.5 systems. See5 (Windows 98/Me/2000/XP) and its Unix counterpart C5.0 are sophisticated data mining tools for discovering patterns that delineate categories, assembling them into classifiers, and using them to make predictions. To maximize interpretability, See5/C5.0 classifiers are expressed as decision trees or sets of if-then rules, forms that are generally easier to understand than neural networks (Quinlan 2000a). Details of the theory and algorithm itself are beyond the scope of this study but reference could be found at (Quinlan 1986).

As explained the rule induction is about generating rules from examples, a process of reasoning from the specific to the general. In this study, the categorisations are straightforward, categorising datasets, whose output falls within a certain pre-defined range, into a group. For the electricity module, for example, categorise the electricity value (the output) into 11 groups, [0, 5), [5, 10), [10, 15), [15, 20), [20, 25), [25, 30), [30, 35), [35, 45), [45, 60), [60, 75), and [75, 100). The grouping did not have same interval but to optimise the size of the group (number of datasets within the group). In the Excel sheet, a number was assigned to each group and presented in a separate column, which would be read as target/output in the rule induction system.

### 5.4.2 Rule Induction Implementation

i) First of all, the database driver should be added to the core library so other software programmes can access the database. This is through ODBC (Open Database Connectivity), which provides a standard software API (Application Programme Interface) method for using database management systems (DBMS). Procedure to do so is as follow: go to control panel – administrative tool – data sources (ODBC) – User DSN add – Microsoft excel driver (.xls) – fill data source name and description, select the path for the targeted data file. For example the Microsoft Excel file for the electricity module was stored in the C drive at

C:\Documents and Settings\Administrator\Desktop\MatlabWork\DataGenerate\elec


**Figure 5-2 ODBC**

ii) Start Clementine.

iii) In Clementine workspace, add database node, open and edit.

a) Drag database node from the source window to the workspace or click data source and add new database.

b) Double click the database node, in the database connection window, connection – select the file added in ODBC, e.g. elec.xls, name appeared in connections frame – click ok. So the Excel file is connected to the Clementine.

c) In database window – in table name, click select.

d) This step selects the specific sheet in Excel that contains the input and output target data. So in Select Table/View window, tick "systems tables" – sheet names appear, select "eleclus" sheet in this case, which contains the inputs and targeted outputs (clusters). In database window, click apply (Figure 5-3).

e) In database window, switch from the default "data" page to "types" page. Assign the types of the parameters to each variable, e.g. employee number as Range and cluster as Set… turn missing as on, read values in (so the data is added in the Clementine besides the structure of the database).

f) Change direction of cluster as out and rest remain in. Therefore inputs and outputs are specified. Figure 5-4.

g) Apply, ok.

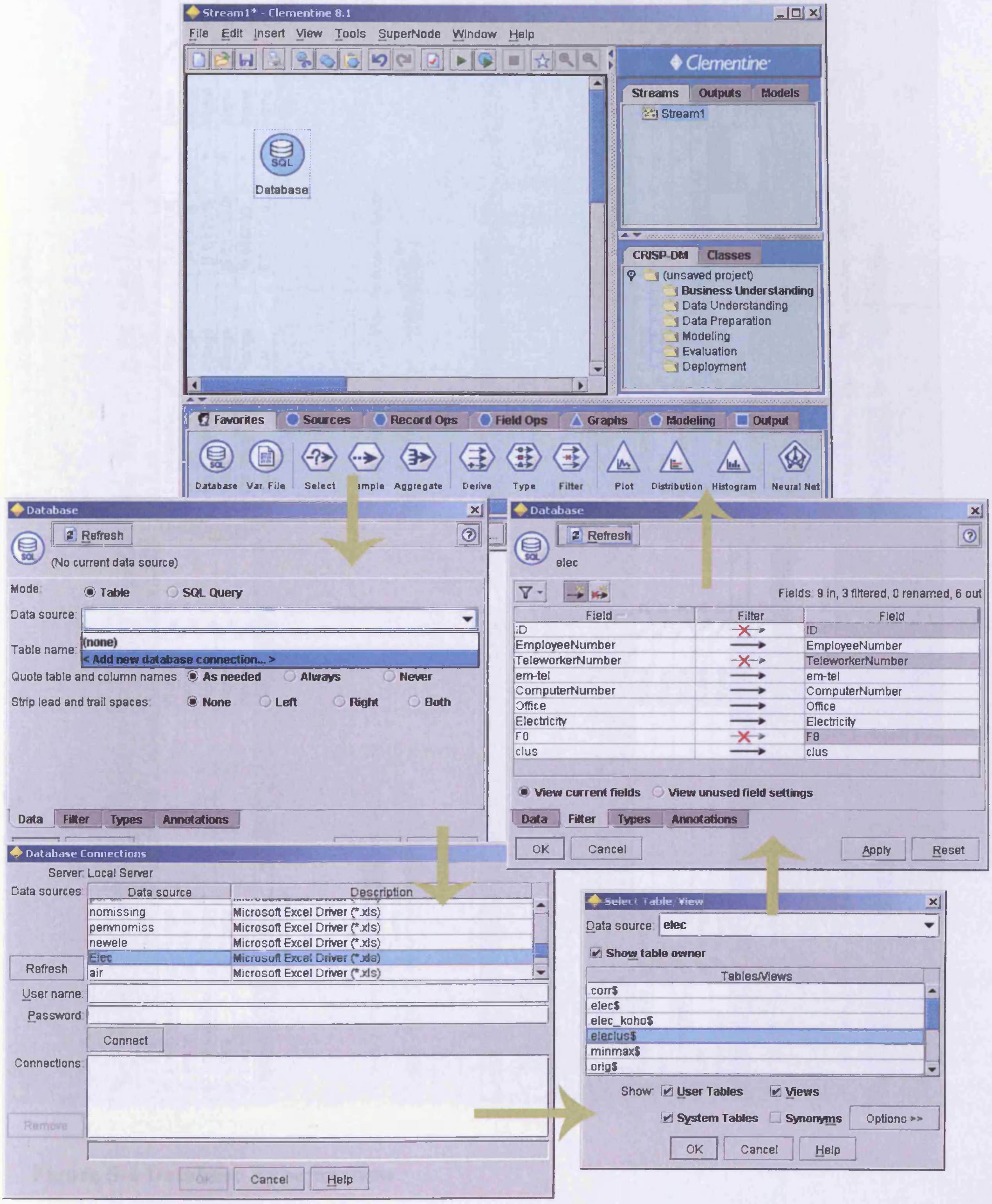



**Figure 5-3 ODBC connection with Clementine**

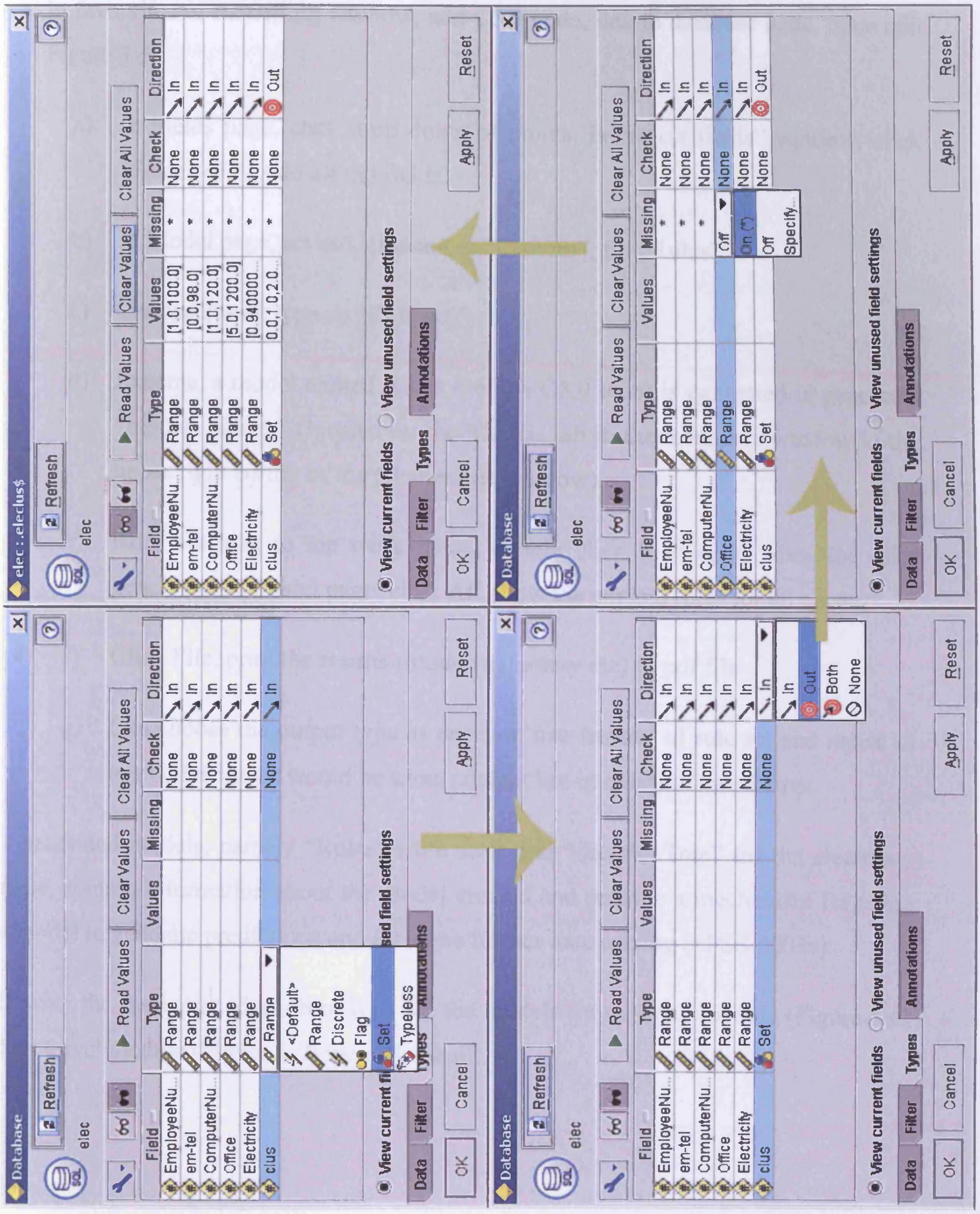



**Figure 5-4 Database Type window**

iv) In favourites or modelling window, add C5.0 node, link to database node, open edit Figure 5-5.

a) In Fields page, click drop down of inputs. In "select fields" window, click "Select All". Add all the fields.

b) In Model page, set model name as "Custom", fill "Rules".

c) Choose output type as "Rule set".

d) Execute, a model named Rules (yellow C5.0 icon) is generated in generated "models palette" (located on the Models tab in the managers window in the upper right corner of the Clementine window).

e) Add this icon to the work place, double click it to open, view the rules generated, in Model page, click All, view the expand rules for all clusters

f) Click File, print the results (model, summary etc) to pdf file

g) Can choose the output type as decision tree instead of rule set and repeat c) to e). The output would be same content but in decision tree format.

The generated models, namely "Rules" (rule sets) and "decisionTree" for the electricity module, contain information about the model created and provide a mechanism for using that model to generate predictions and facilitate further data mining (SPSS 2003a).

Following the same procedure from i) to iv), the models for the road module (Figure 5-6) and air travel module (Figure 5-7) were also built.

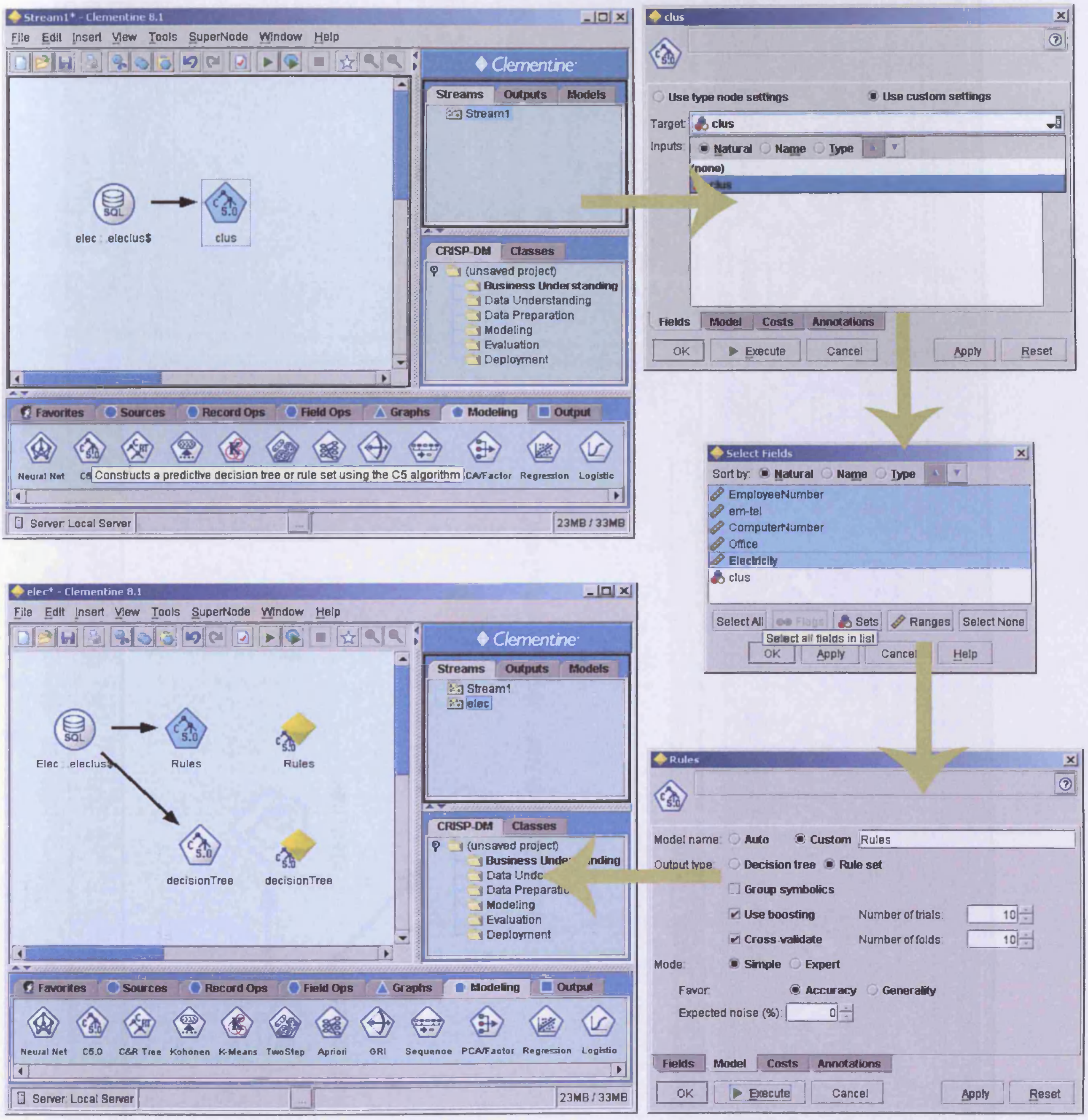



**Figure 5-5 C5.0**

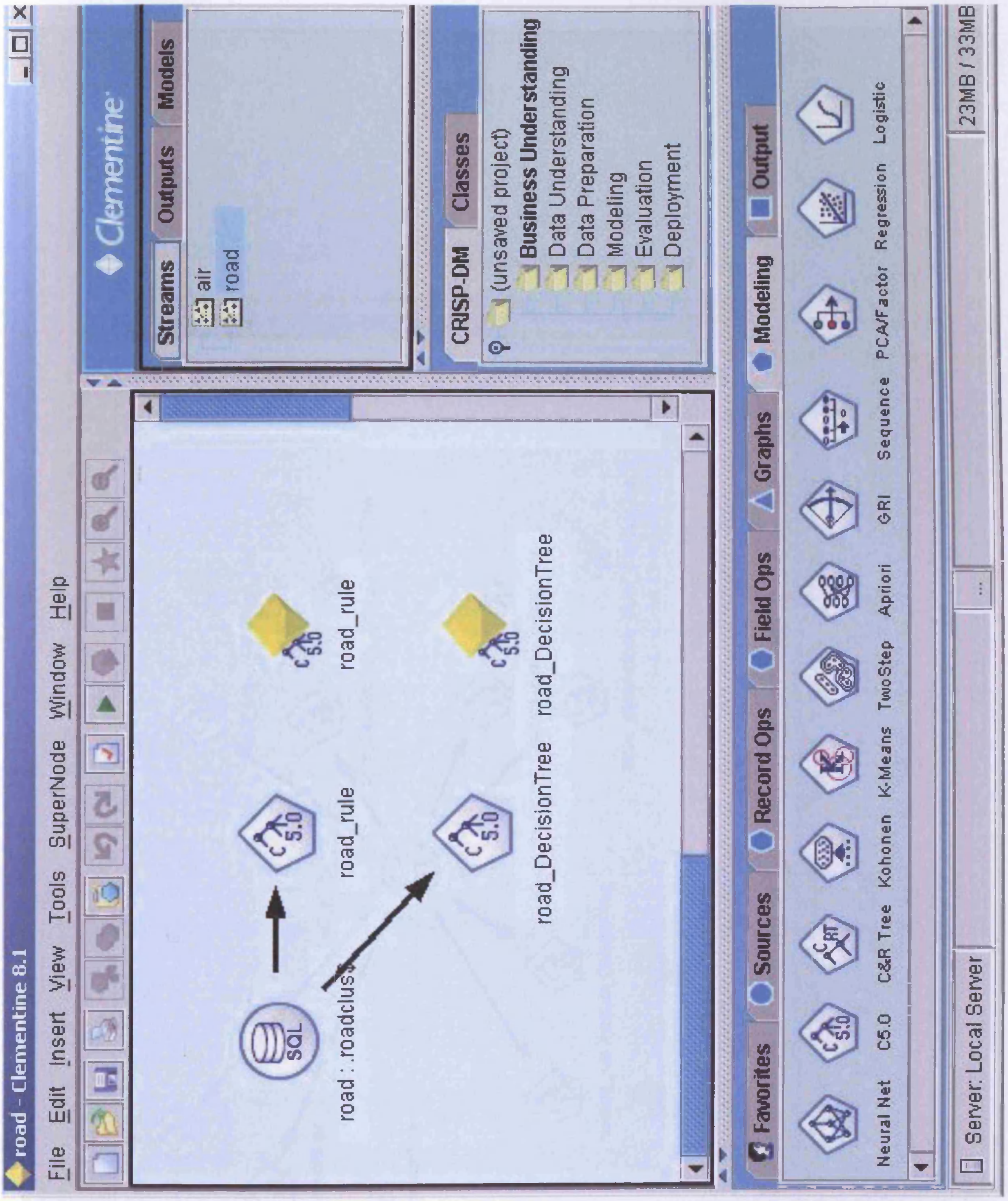



**Figure 5-6 Road module rule induction**

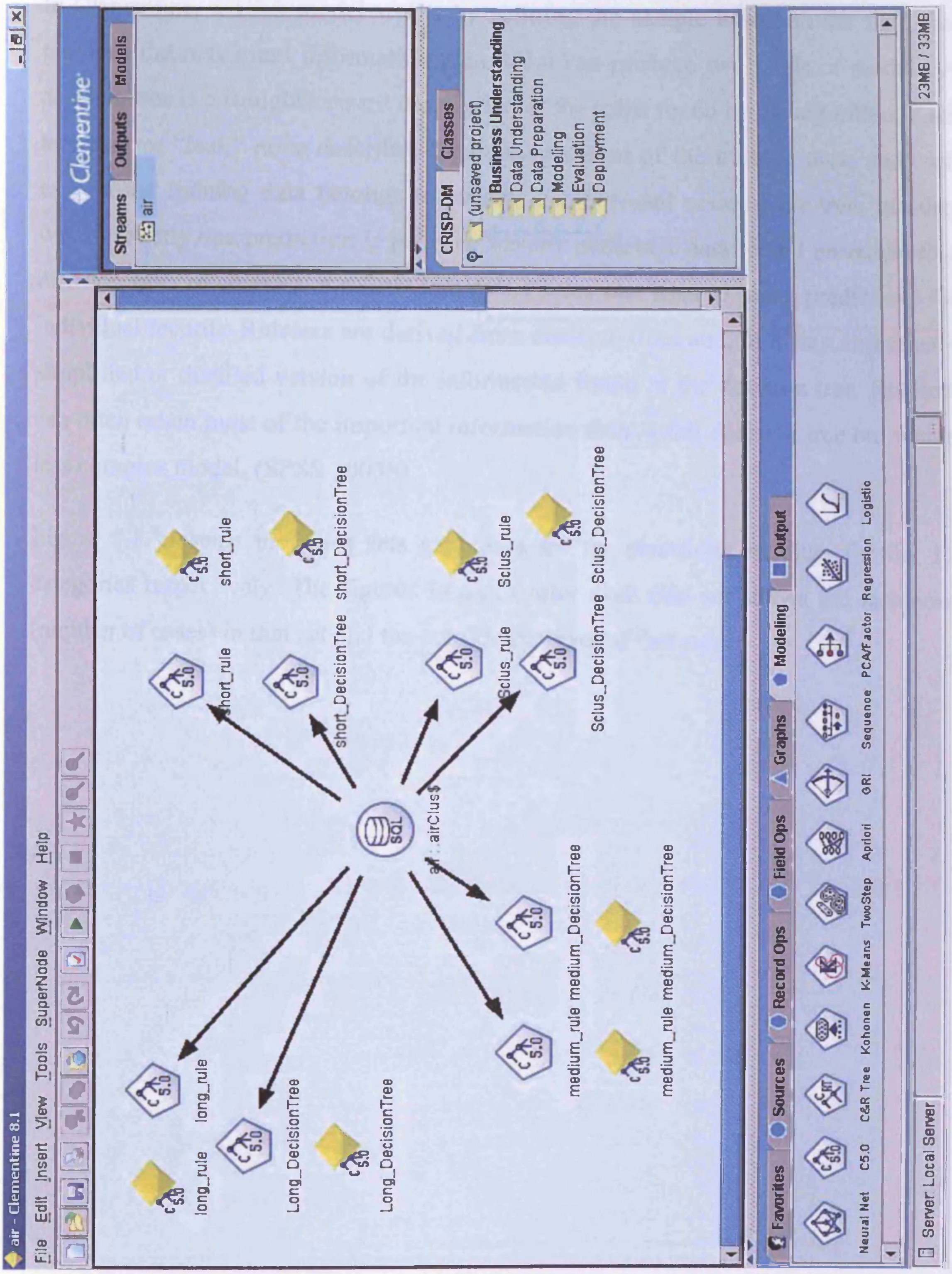



**Figure 5-7 Air travel module rule induction**

### 5.4.3 Results: Decision Trees and If-then Rule Sets

In Clementine, a C5.0 model works by splitting the sample based on the field that provides the maximum information gain. C5.0 can produce two kinds of models. A decision tree is a straightforward description of the splits found by the algorithm. Each terminal, or "leaf," node describes a particular subset of the training data, and each case in the training data belongs to exactly one terminal node in the tree. In other words, exactly one prediction is possible for any particular data record presented to a decision tree. In contrast, a ruleset is a set of rules that tries to make predictions for individual records. Rulesets are derived from decision trees and, in a way, represent a simplified or distilled version of the information found in the decision tree. Rulesets can often retain most of the important information from a full decision tree but with a less complex model. (SPSS 2003b)

Figure 5-8 presents the rules sets generated for the electricity module, for the 11 categories respectively. The figures in green after each rule set shows the instances (number of cases) in that set and the confidence level of that rule.

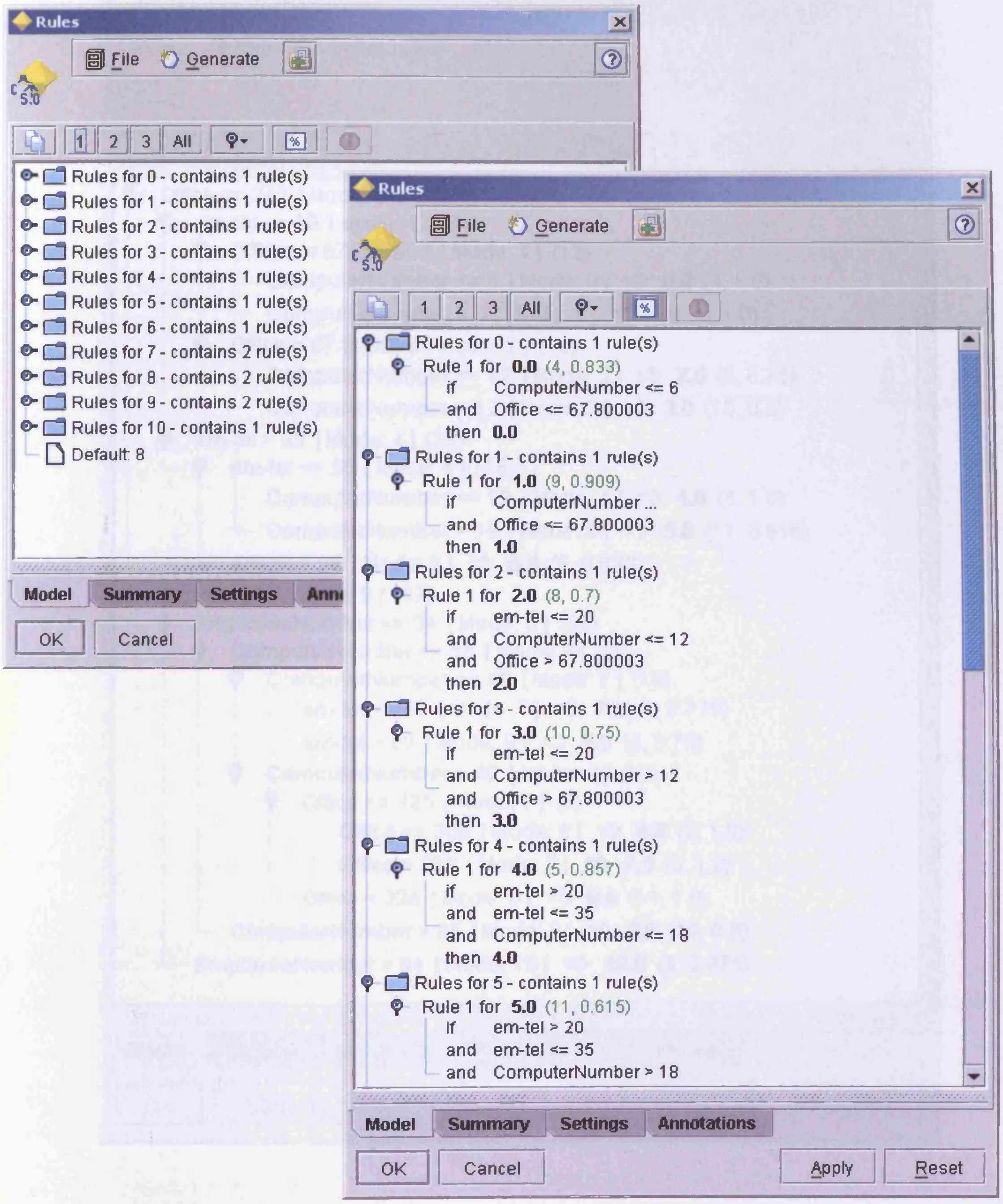



**Figure 5-8 Electricity module rule sets**

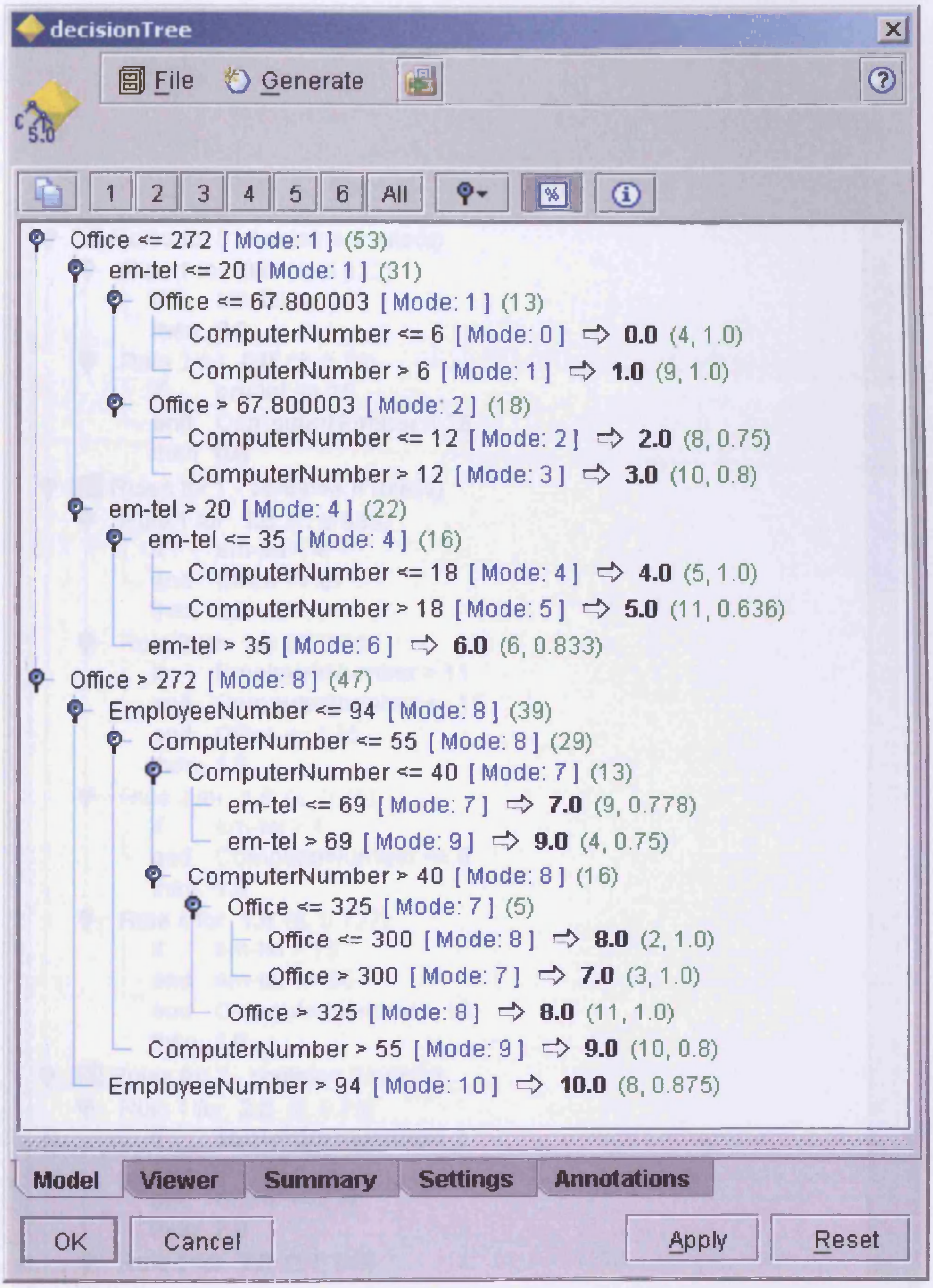



**Figure 5-9 Electricity module decision tree**

In the viewer window, the decision tree can be viewed in different graphical formats, including "frequency information" in a table and as a graph; the decision tree in top-down orientation, left to right, and right to left orientation. The graphics in Clementine are particularly user-friendly.

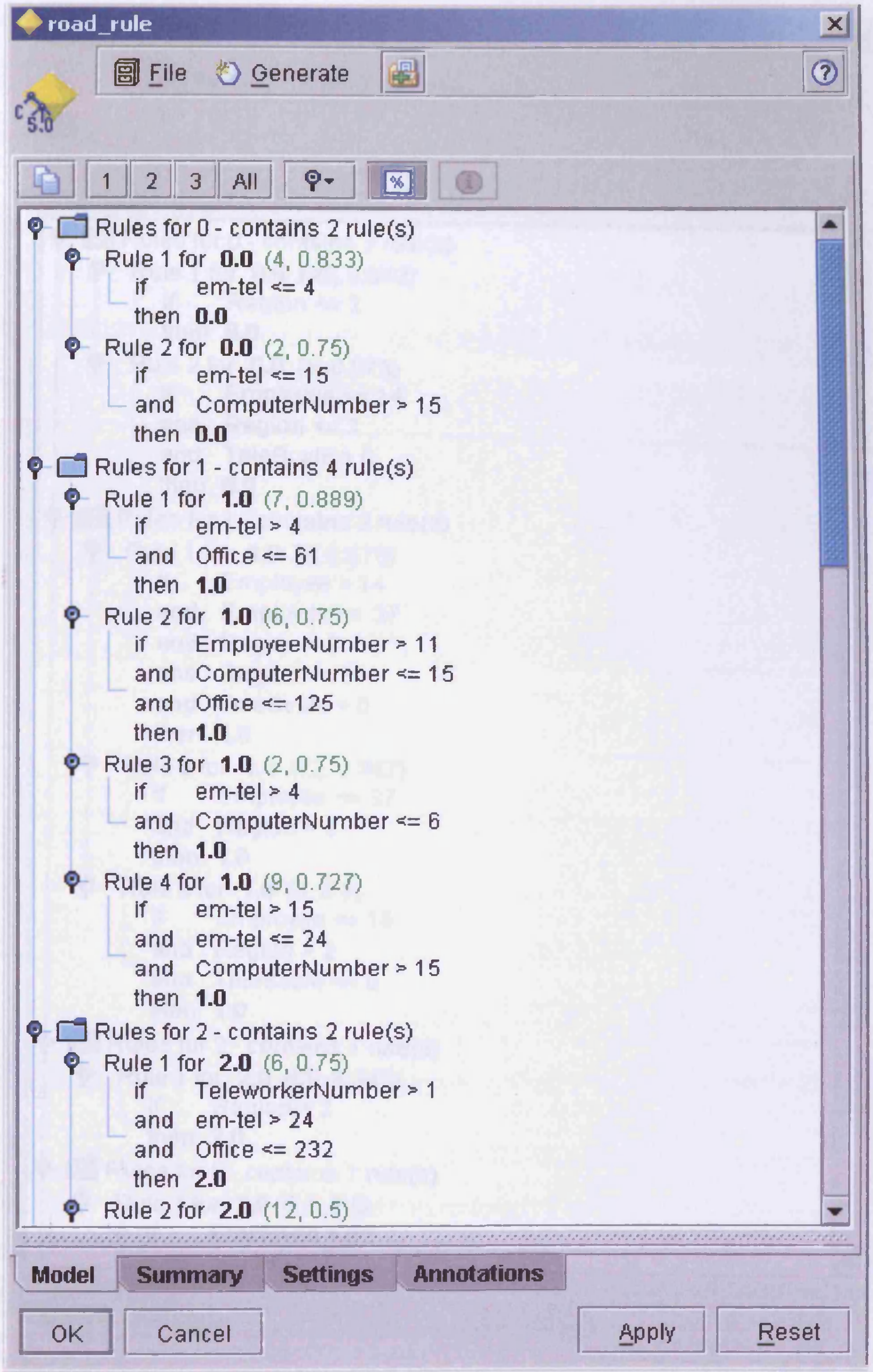



**Figure 5-10 Road module rule sets**

Figure 5-8 presents the rule sets generated for the road module and Figure 5-11 shows the rule set for the medium-haul of air travel module.

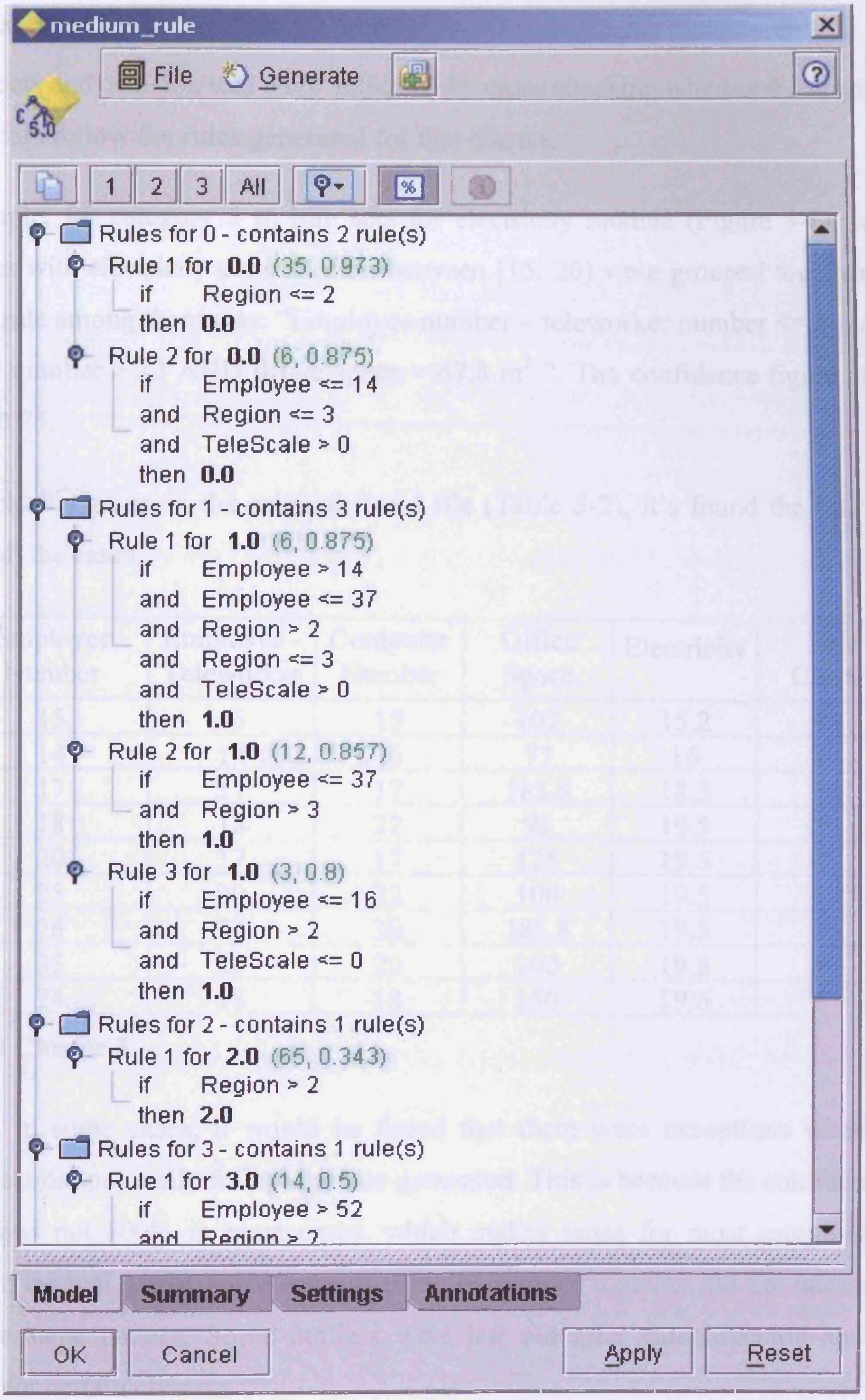



**Figure 5-11 Medium haul air travel rule sets**

### 5.4.4 Validation

The rule sets and decision tree were validated by cross checking whether the originally grouped data follow the rules generated for that cluster.

For example, for category 3 in rule sets for electricity module (Figure 5-8), where companies with electricity consumption between [15, 20) were grouped together, the common rule among them was: "Employee number – teleworker number <= 20, AND computer number > 12 AND office space > 67.8 $m^2$ ". The confidence figure of this rule was 0.75.

Trace back this group in the original Excel file (Table 5-2), it's found the rule does apply to all the cases.

| ID | Employee Number | Employee - Teleworker | Computer Number | Office Space | Electricity | Elec Cluster |
|---|---|---|---|---|---|---|
| 17 | 15 | 15 | 15 | 102 | 15.2 | 3 |
| 16 | 14 | 14 | 16 | 77 | 16 | 3 |
| 22 | 17 | 17 | 17 | 185.5 | 18.5 | 3 |
| 24 | 18 | 18 | 22 | 92 | 19.5 | 3 |
| 27 | 20 | 17 | 17 | 125 | 19.5 | 3 |
| 31 | 25 | 20 | 22 | 100 | 19.5 | 3 |
| 35 | 26 | 24 | 30 | 185.8 | 19.5 | 3 |
| 45 | 35 | 20 | 20 | 100 | 19.5 | 3 |
| 29 | 24 | 18 | 18 | 150 | 19.6 | 3 |

**Table 5-2 Cluster 3**

However, in some cases, it would be found that there were exceptions where the original data did not really follow the rule generated. This is because the confidence of the rule was not 100% in most cases, which makes sense for most categorisation problem in the real world, since datasets initially grouped together did not necessarily follow the same pattern. Some outliers were left out after categorisation and rule induction for certain clusters.

## 5.5 EDSS Implementation

This section will explain how the EDSS was built based on the trained and ready-to-use neural network and generated rule set models, the Matlab programming.

### 5.5.1 Forward Chaining DSS

In the previous chapter, section 4.6.4 has given an example of the network structure of the road module. This result provides all the parameter values needed to build the DSS. Three M files (.m extension) were programmed for the three neural nets respectively. M-files can be scripts that simply execute a series of MATLAB statements, or they can be functions that also accept input arguments and produce output (MathWorks 1984-2001). In this case, where inputs arguments to be read and outputs arguments to be generated, the M files were written in the function format.

A typical M files contains:

- 'Function definition line' defining the function name, and the number and order of input and output arguments
- H1 line: A one line summary description of the program
- Help text: A more detailed description of the program
- Function body: Program code that performs the actual computations and assigns values to any output arguments
- Comments: Text in the body of the program that explains the internal workings of the program

The function definition line and the function body are the essential components for the primary function. The function name is to be called by other functions outside of the M files and be invoked by Matlab command line.

The following presents the air function, with three input arguments *input1, input2* and *input3,* and three output arguments, *airlong, airmed* and *airshort.*

```
function [airlong,airmed,airshort] = air(input1,input2,input3)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%% Customer Network Air Module %%%%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

net_air = network(1,2,[1; 1], [1; 0],[0 0; 1 0],[0 1],[0 1]);

net_air.adaptFcn = 'trains';

net_air.initFcn = 'initlay';

net_air.performFcn = 'mse';

net_air.trainFcn = 'trainlm';


net_air.inputs{1}.size = 3; %%%% 3 variables in input layer

net_air.numLayers = 2;     %%%% 2 hidden layer, one tansig one purelin

net_air.layers{1}.size = 9; %%%% 1st hidden layer with 9 neurons

net_air.layers{1}.transferFcn = 'tansig';

net_air.layers{2}.transferFcn = 'purelin';

net_air.layers{2}.size = 3;

net_air.outputs{1}.size = 3;

net_air.targets{1}.size = 3;

%%%%minmax

net_air.inputs{1}.range = [-1 1; -1 1; -1 1];

%%%%% Inputs weights IW{i,j} ith layer, jth input
```

```
net_air.IW{1,1} = [...
   32.5535  -60.2860  -40.1334;
    0.6287   12.0599   -0.3022;
   26.5155  -74.6113  -14.3825;
    0.6409    1.7352   -0.3097;
    0.5551   13.3855   -0.1714;
   -0.5327   -7.8785    0.1259;
   -1.2382  -13.1309    2.7549;
    0.9051   18.8409   -1.7995;
    2.1733   -3.1244   -0.1435];
net_air.IW{2,1} = [];
%%%%% bias
net_air.b{1} = [2.2785;-11.9773;-4.1828;-1.6717;...
   -5.8547;4.2044;15.6061;-16.9116;3.4383];
net_air.b{2} = [0.4286;-1.5594;-0.7596];
%%%%%% Layer weights
net_air.LW{1,1} = [];
net_air.LW{1,2} = [];
net_air.LW{2,2} = [];
net_air.LW{2,1} = [...
   -0.0177   1.6128   0.0362   0.0742   0.9055   1.1114  -0.3265  -0.5408  -0.1845;
```

```
0.2620  -1.9767  -0.4784  -0.0107  11.0146  10.8305  -0.5022  1.5663  1.0432;

-0.0339  -14.8263  -0.0460  14.5885  -81.2593  -81.6401  -0.0848  0.1381
0.1662];

%%%%%% minmax inputs targets for pre- and post- processing

minp_air = [1;1;0];

maxp_air = [100;4;5];

mint_air = [0;0;0];

maxt_air = [5;4;20];


test = [input1;input2;input3];

testn = tramnmx(test, minp_air, maxp_air);

testn = sim(net_air, testn);

out = postmnmx(testn,mint_air,maxt_air);

airlong = round(out(1));

airmed = round(out(2));

airshort = round(out(3));
```

Basically this function defined the network structure (size, transfer functions of each layer, initial weights, bias, layer bias, layer weights and so on) and created a neural network for the air travel module, and is able to take three input arguments, process them in this modular network and produce three output arguments, which are the predicted long, medium and short haul air travel frequencies.

Similarly two M files/functions, road.m and elec.m (both with 4 input arguments and 1 output argument), were programmed for the road and electricity module as follow. When one of these M-file functions is called, MATLAB parses and executes each line of code in the M-file. It saves the parsed version of the function in memory, eliminating parsing time on any further calls to this function (MathWorks 1984-2001).

```
=====================================================

function roadout = road(input1,input2,input3,input4)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%%% Customer Road Module %%%%%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

net_road = network(1,2,[1; 1], [1; 0],[0 0; 1 0],[0 1],[0 1]);

net_road.adaptFcn = 'trains';

net_road.initFcn = 'initlay';

net_road.performFcn = 'mse';

net_road.trainFcn = 'trainlm';

net_road.inputs{1}.size = 4; %%%% 4 variables in input layer


net_road.numLayers = 2;     %%%% 2 hidden layer, one logsig one purelin

net_road.layers{1}.size = 4; %%%% 1st hidden layer with 4 neurons
```

```
net_road.layers{1}.transferFcn = 'logsig';

net_road.layers{2}.transferFcn = 'purelin';

%%%%minmax

net_road.inputs{1}.range = [-1 1; -1 1; -1 1; -1 1];

%%%%% Inputs weights IW{i,j}ith layer, jth input

net_road.IW{1,1} = [...

  -45.5242   81.4700    2.8699  -54.8121;

  -45.9917   21.4087   -9.9876  -41.8836;

    0.3607   -1.2531    0.2346   -0.8943;

    2.0671    4.3992   24.8885   33.6262];

net_road.IW{2,1} = [];

%%%%%  bias

net_road.b{1} = [15.6411; 9.6372; -1.5016; -48.6610];

net_road.b{2} = 0.0143;

%%%%%% Layer weights

net_road.LW{1,1} = [];

net_road.LW{1,2} = [];

net_road.LW{2,2} = [];

net_road.LW{2,1} = [0.4838,-0.4001,-2.0761,4.5656];

%%%%%% minmax inputs targets for pre- and post- processing
```

```
minp_road = [1;0;1;5];

maxp_road = [100;98;120;1200];

mint_road =  0;

maxt_road =  15000;


test = [input1;input2;input3;input4];

testn = tramnmx(test, minp_road, maxp_road);

testn = sim(net_road, testn);

roadout = postmnmx(testn,mint_road,maxt_road);
```

---

```
function elecout = elec(input1,input2,input3,input4)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%%% Customer Network Electricity Module %%%%%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

net_elec = network(1,2,[1; 1], [1; 0],[0 0; 1 0],[0 1],[0 1]);

net_elec.adaptFcn = 'trains';

net_elec.initFcn = 'initlay';

net_elec.performFcn = 'mse';

net_elec.trainFcn = 'trainlm';

net_elec.inputs{1}.size = 4;
```

```
net_elec.numLayers = 2;

net_elec.layers{1}.size = 3;


net_elec.layers{1}.transferFcn = 'tansig';

net_elec.layers{2}.transferFcn = 'purelin';


net_elec.inputs{1}.range = [-1 1; -1 1; -1 1; -1 1];


net_elec.IW{1,1} = [...
  23.6952   19.1270   11.8550   -0.1417;
  -0.7609    0.5101   -2.0576   -1.7941;
   0.0101   -0.0979   -0.1078   -0.0588];

net_elec.IW{2,1} = [];

%%%%%%%%%%%

net_elec.b{1} = [-44.5392; 0.2936; -1.7734];

net_elec.b{2} = -28.9079;

%%%%%%%%%

net_elec.LW{1,1} = [];

net_elec.LW{1,2} = [];

net_elec.LW{2,2} = [];

net_elec.LW{2,1} = [-0.7483   -0.2623  -30.1628];
```

```
minp_elec = [1;0;1;5];

maxp_elec = [100;98;120;1200];

mint_elec = 0.9400;

maxt_elec = 94;


test = [input1;input2;input3;input4];

testn = tramnmx(test, minp_elec, maxp_elec);

testn = sim(net_elec, testn);

elecout = postmnmx(testn,mint_elec,maxt_elec);
```

### 5.5.2 Backward Chaining ES

The backward chaining ES system is basically to transfer the rule sets generated in Clementine (in Clementine node or printed HTML or PDF format) and to code them in Matlab environment, so corresponding rule set to be triggered, invoked and presented once certain inputs were presented. This part of programming was integrated with the Graphical User Interface to be explained in next section, so the implementation will be explained together with the GUI development.

## *5.6 The Graphical User Interface*

### 5.6.1 The Concept and the Output

The programme of the EDSS, which was built in Matlab, involves understanding of Matlab commands and programming. The end-users would not be expected to understand how the functions, neural nets, and all the background programming work to be able to operate the EDSS. Learning, remembering, and using many different keyboard commands can be intimidating. To provide an easier method communicating with the operating system, a graphical user interface (GUI) is thus needed. A GUI uses dialog boxes, drop-down and pop-up menus, buttons, icons, scroll bars, pointers and other devices instead of requiring commands (Schultheis and Summer 1998).

In MATLAB, the Graphical User Interface development environment is called GUIDE, which provides a set of tools for building and programming GUIs and greatly simplifies the process of designing and building GUIs.

Figure 5-12 presents the final GUI built for the EDSS system, with two sections in Figure 5-13 interchangeable with the Environmental Indicators Prediction part.

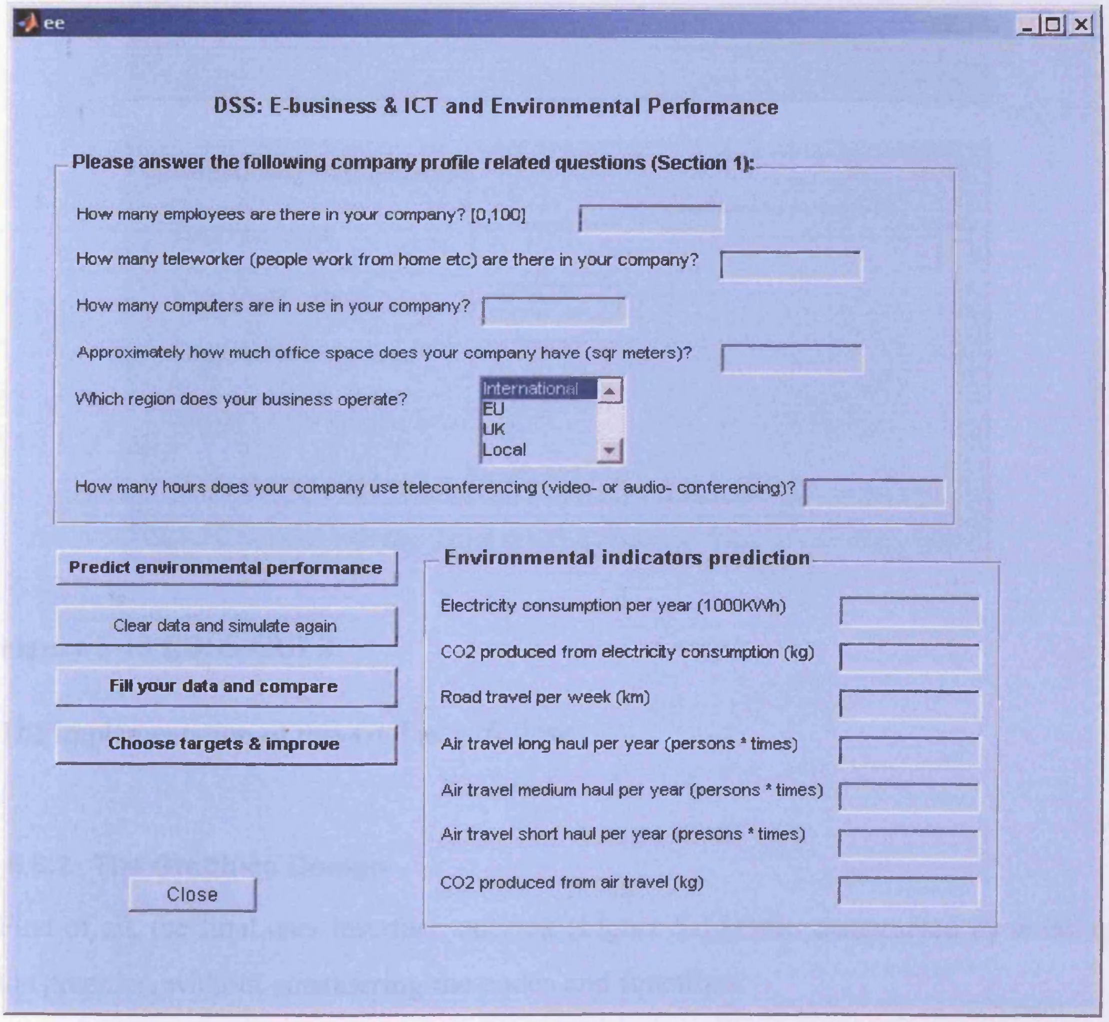



**Figure 5-12 EDSS GUI**

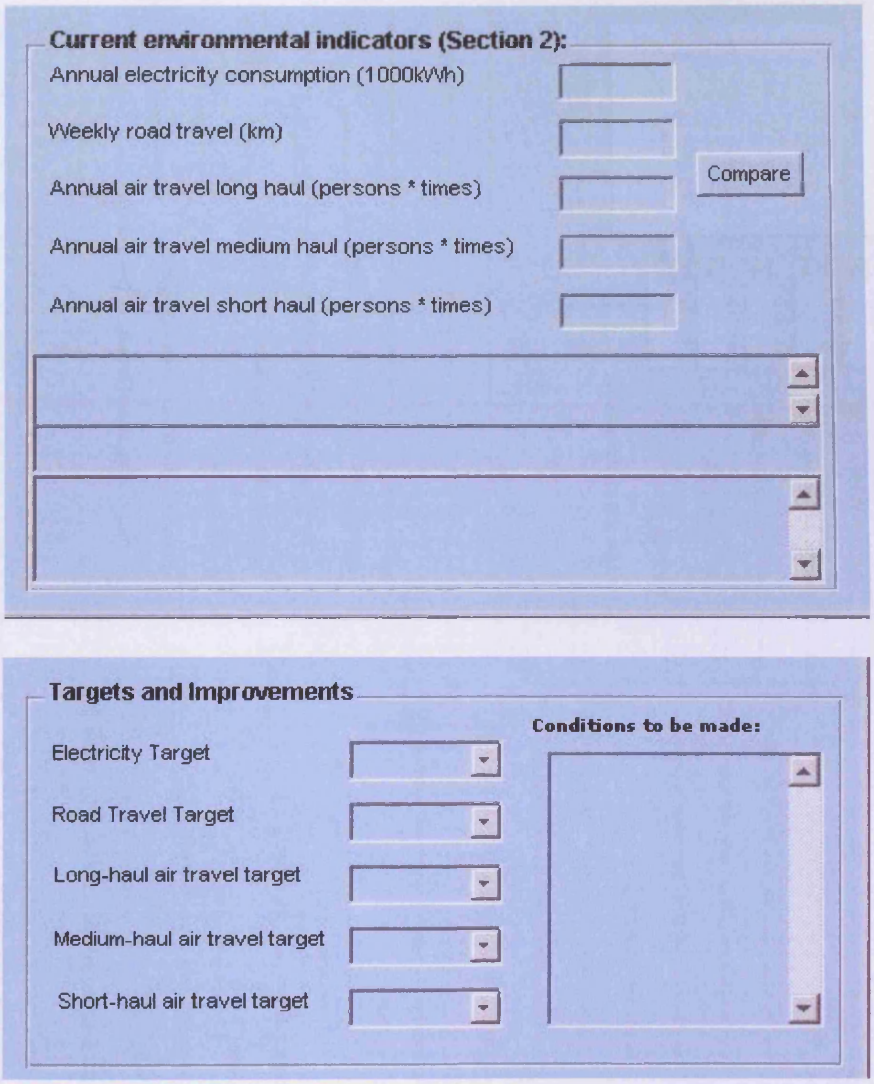



**Figure 5-13 EDSS GUI 2**

The implementation of this GUI is as follow.

### 5.6.2 The Graphics Design

First of all, the final user interface outlook (Figure 5-14) was constructed by creating the graphics, without considering the codes and functions.

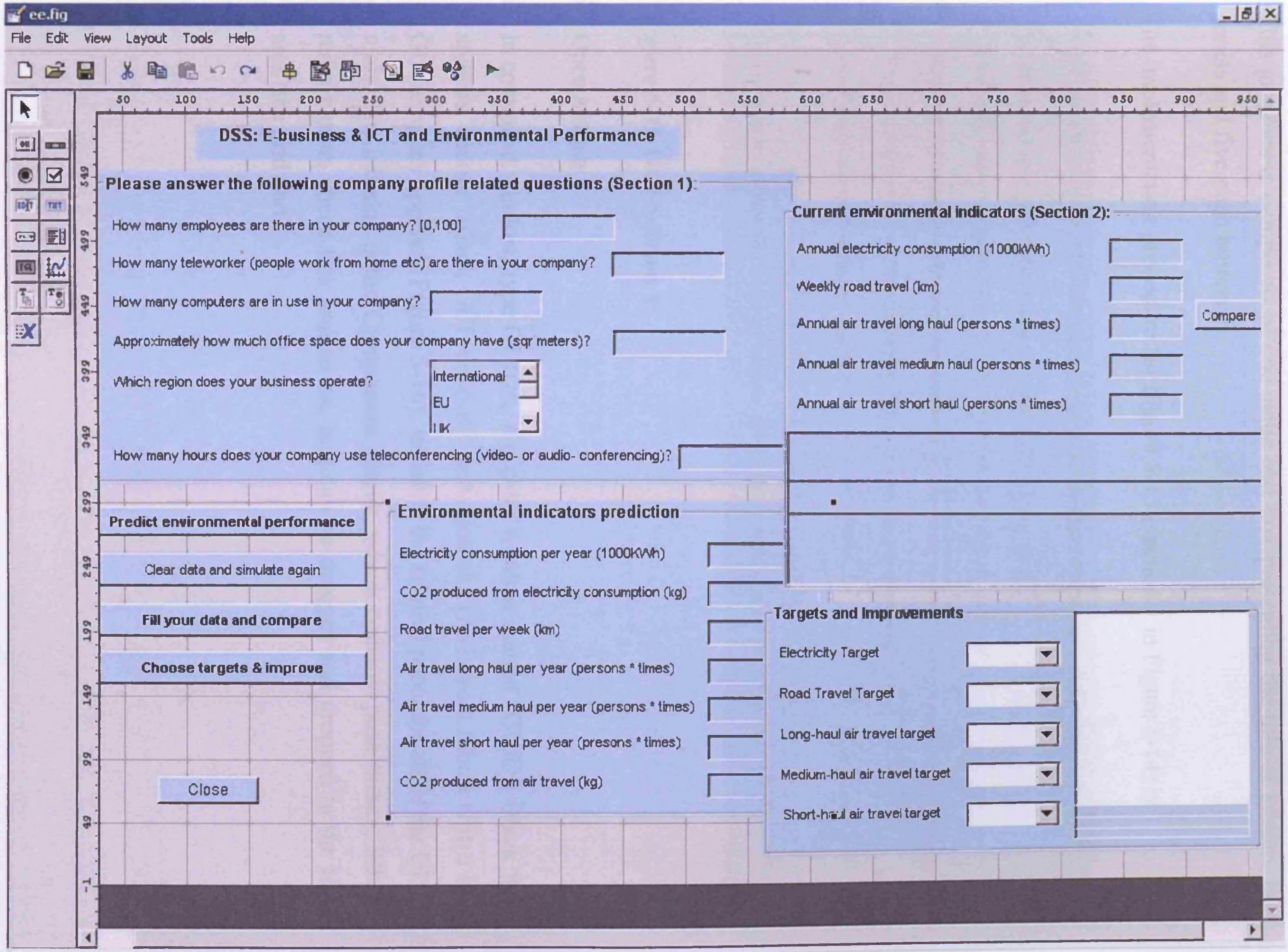



Figure 5-14 GUI Graphical Design

This graphics is composed of one static text box presenting the title of system, four panels and five push buttons.

The implementation procedure for Figure 5-15 (section 1 in Figure 5-14) is:

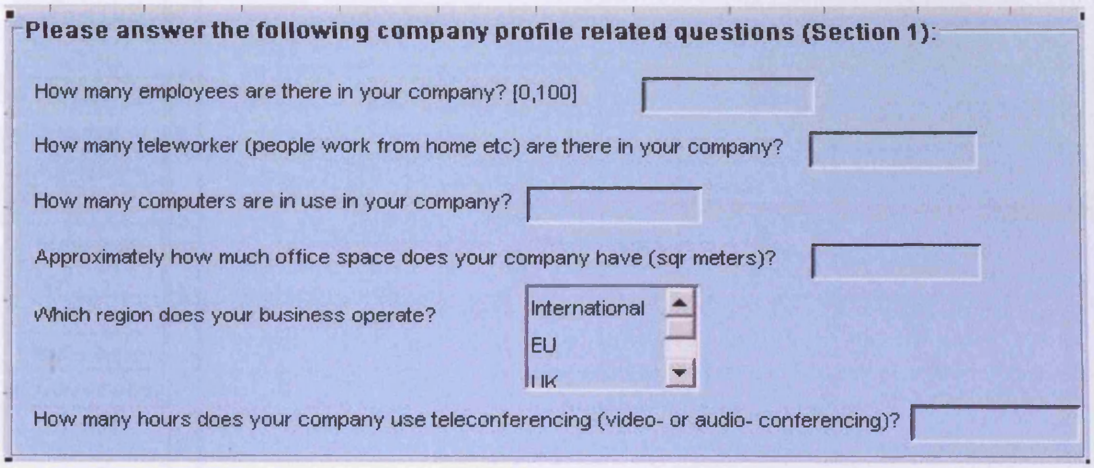



**Figure 5-15 GUI Section 1**

i) Open Matlab.

ii) In command window, type GUIDE, a pop-up window named "GUIDE Quick Start" appears, choose "Blank GUI (default)" and click ok. The Layout Editor with a blank GUI template appears (Figure 5-16), which is the control panel for all of the GUIDE tools. The left side is the Component Palette, which contains push button, list box, radio button, check box and so on, and can be dragged and dropped in the layout area (the grid part).

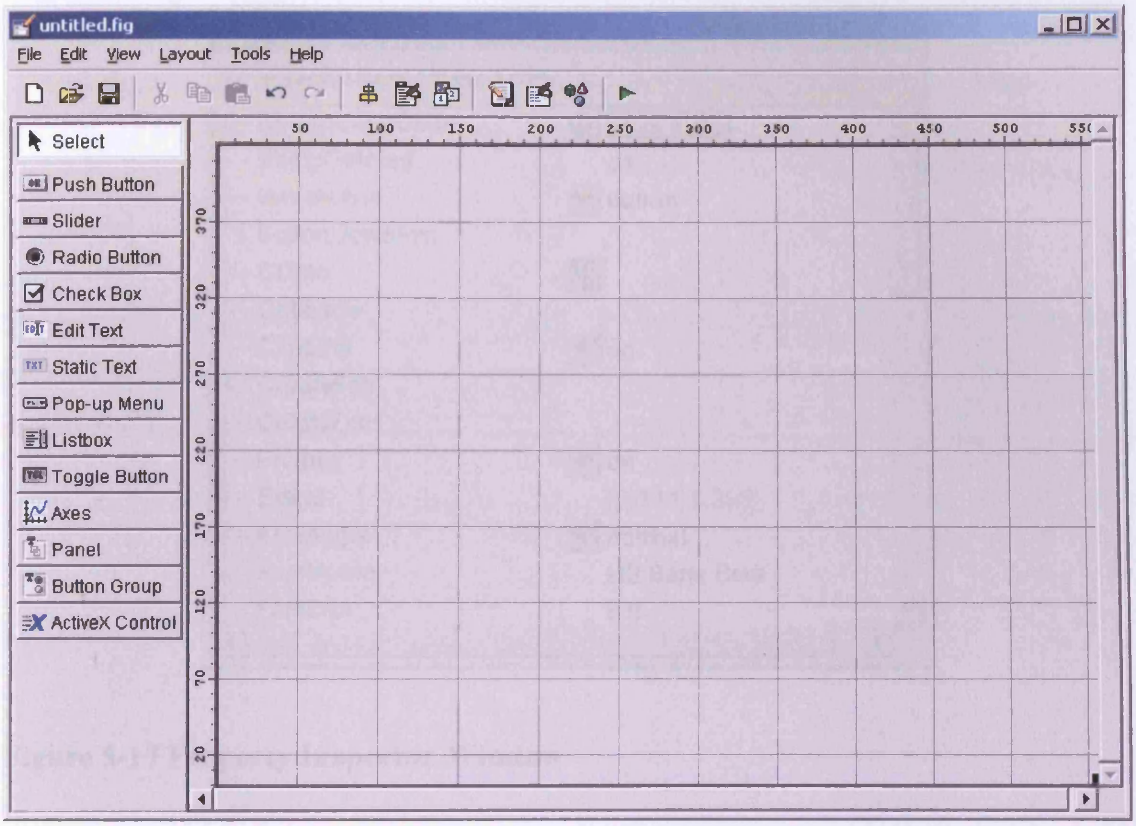



**Figure 5-16 The Layout Editor**

iii) Drag and drop the static text box in the layout area, and double click it, the property window (Figure 5-17) appears. Scroll down and fill "DSS: E-business & ICT and Environmental Performance" in the "String" field, which defines the title of this EDSS.

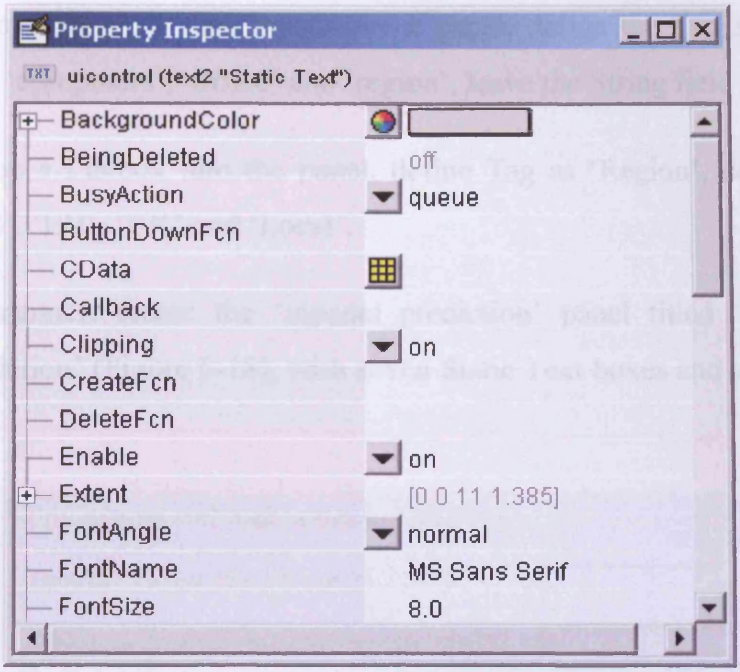



**Figure 5-17 Property Inspector Window**

iv) Drag and drop the panel in the layout area, double click it to open the Property Inspector of the panel, define Tag as 'uipanel-companyprofile', Title as 'Please answer the following company profile related questions (Section 1):'. The Tag property provides a string as a unique identifier for each component. GUIDE uses this identifier to construct unique callback names for the different components in the GUI. The String property is to set the label in some user interface controls.

v) Within the panel, drag and drop a static text box inside, define Tag as 'employee-label', String as 'How many employees are there in your company? [0,100]'.

vi) Similarly drag and drop five static text boxes, define the Tag as 'teleworker-label', ComputerNumber', 'OfficeSpace', 'region-label' and 'teleconf-label' respectively, and String as 'How many teleworker (people work from home etc) are there in your company?', 'How many computers are in use in your company?', 'Approximately how much office space does your company have (sqr meters)?', 'Which region does your business operate?', and 'How many hours does your company use teleconferencing (video- or audio- conferencing)?' correspondingly.

vii) Drag and drop four Edit Text boxes in the panel, define the Tag as 'Employee', 'teleworker', 'computers', 'office' and 'region', leave the String field as blank.

viii) Drag and drop a Listbox into the panel, define Tag as 'Region', define String as 'International', 'EU', 'UK' and 'Local'.

In a similar manner, create the 'uipanel_prediction' panel titled 'Environmental indicators prediction' (Figure 5-18), with seven Static Text boxes and seven Edit Text boxes.

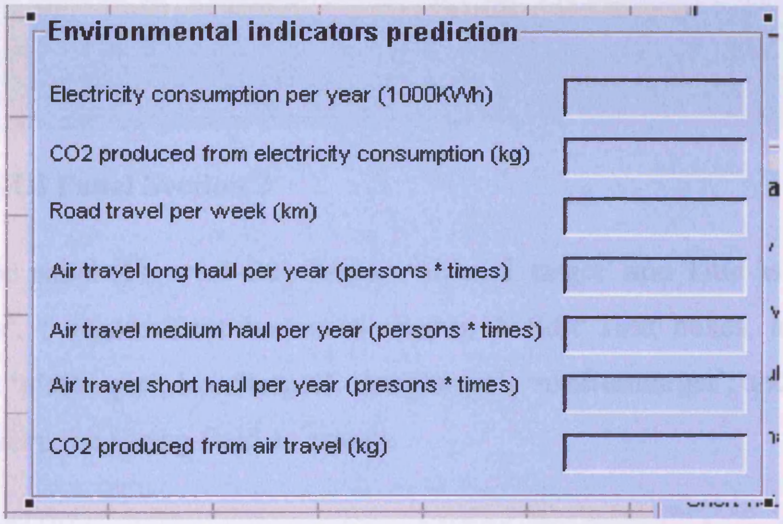



**Figure 5-18 GUI Panel 'Environmental indicators prediction'**

Also create the 'uipanel_currentEnv' Panel titled 'Current environmental indicators (Section 2):' (Figure 5-19), with 5 Static Text boxes, 8 Edit Text boxes, and one Push button (Tag as 'comparebutton', String as 'Compare').

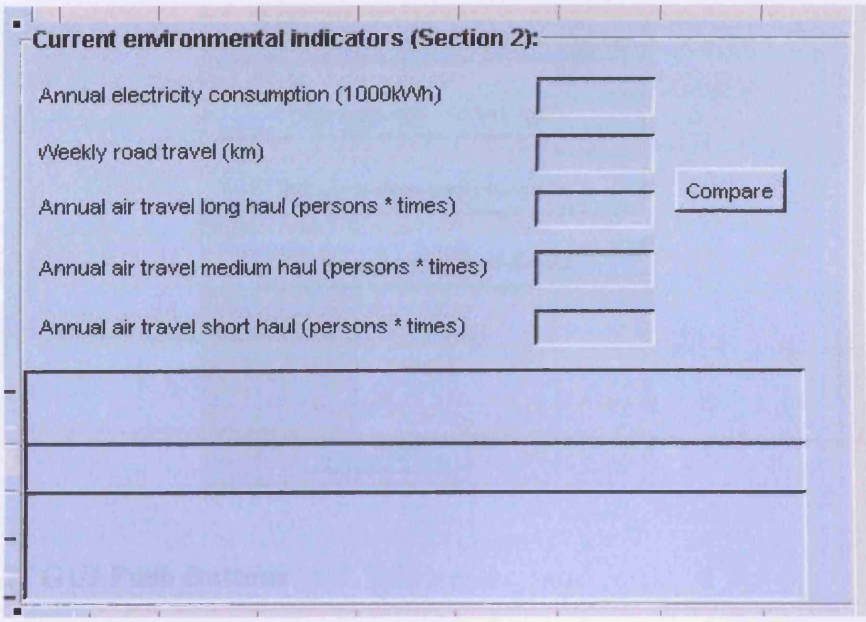



**Figure 5-19 GUI Panel Section 2**

Also create the panel (Figure 5-20) Tag as 'uipanel_target' and Title as 'Targets and Improvements', 5 Static Text boxes as shown, 5 Edit Text boxes, and 5 Pop-up Menus Tag as 'electarget', 'roadtarget', 'longtarget', 'mediumtarget', and 'shorttarget' respectively, keep the String field as empty.

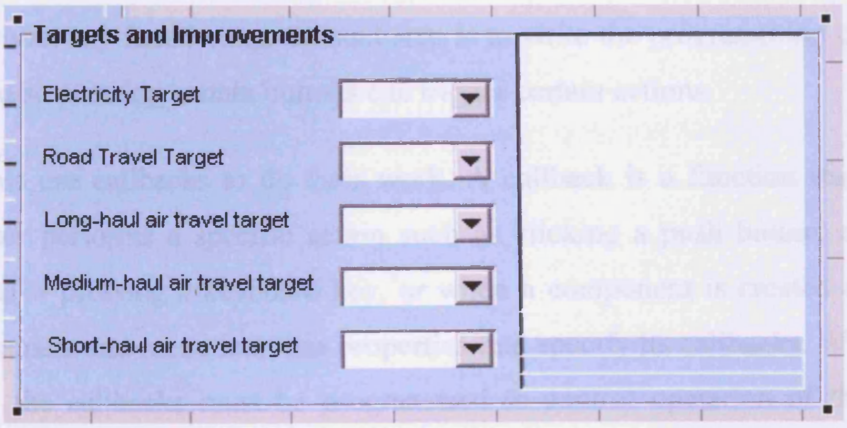



**Figure 5-20 GUI Panel 'Target and Improvements'**

Finally, drag and drop 5 push buttons, define the tag as 'simbutton', 'clearbutton', 'fillenvdata', 'targetbutton' and 'closebutton' respectively, fill the String as shown in Figure 5-21.

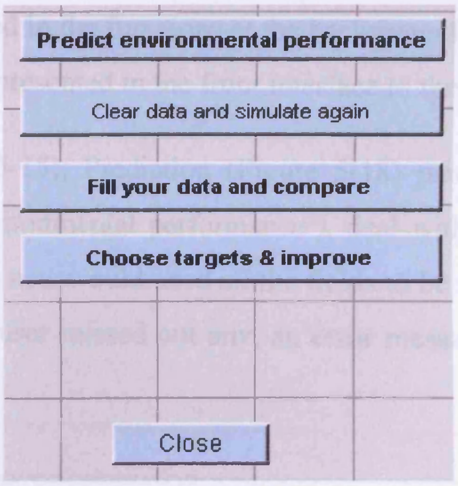



**Figure 5-21 GUI Push Buttons**

The graphical design part is completed until now. The layout should look like Figure 5-14 more or less (the background colour is defined as Red: 187, Green: 205, Blue: 223).

### 5.6.3 The Callbacks and Functions

So far the GUI, consisting of text boxes, push buttons, list boxes etc, is still a static picture without any functioning. So next step is to write the programming code in the background so pressing certain buttons can trigger certain actions.

Components use callbacks to do their work. A callback is a function that executes when a user performs a specific action such as clicking a push button, selecting a menu item, or pressing a keyboard key, or when a component is created or deleted. Each component and menu item has properties that specify its callbacks. When a GUI is created, the callbacks must be programmed to control operation of the GUI. A component can have many callback properties, but the most common one is the Callback property. The code provide for the Callback property performs the primary work of the component. It executes, for example, when a user presses a push button, moves a slider, or selects a menu item (MathWorks 1984-2001).

These callbacks are basically for linking certain attributes of these graphical components to the input and output arguments of the functions. So operating the front interface graphical components, data can be sent to corresponding field in the codes, be calculated and processed in the functions at the background, and results can be sent back to relevant field and presented in the front interface to the users.

Panel section 1 (Figure 5-15), Prediction (Figure 5-18) panel, and the simulation button (titled 'Predict environmental performance'), deal with the 3 modular neural networks. Since the neural nets would need all the fields to be filled before data could be processed, in case the user missed out any, an error message would pop up. The codes are as follow:

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

```
set(handles.uipanel_prediction,'Visible','on')
set(handles.uipanel_currentEnv,'Visible','off')
set(handles.uipanel_target,'Visible','off')
if isempty(get(handles.Employee,'String'))|...
    isempty(get(handles.teleworker,'String')) |...
  isempty(get(handles.computers,'String'))|...
  isempty(get(handles.office,'String'))|...
  isempty(get(handles.teleconf,'String')),
  errordlg('Please fill all fields in Section 1.')
```

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

The following codes pass the user input of Employee number, teleworker number etc filled in the Edit Text boxes in Panel Section 1 to the function elec.m, corresponding to the 3 input arguments. The function is invoked, executed and output argument is passed back to the Prediction Panel's electricity Edit Text box field. During the process, data type should be converted so the function takes numerical data and the GUI presents string. The $CO_2$ field is calculated from the output electricity.

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

```
elecout = elec(str2double(get(handles.Employee,'String')),...
   (str2double(get(handles.Employee,'String'))-
str2double(get(handles.teleworker,'String'))),...
   str2double(get(handles.computers,'String')),...
   str2double(get(handles.office,'String')));
% %show result in electricity field
set(handles.electricity,'String',num2str(elecout))
% %%%%%%%Grid electricity 1kWh  * 0.43 = kg Co2
set(handles.co2elec,'String',430*str2double(get(handles.electricity,'String')))
```

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

Similar codes were programmed for the road and air module. The Clear button resets all the fields in these two panels, for example with the following code for Employee field.

```
set(handles.Employee,'String',[],'Value',0)
```

All the panels except the section 1 (Figure 5-15) are interchangeable, which means only one would be active and presented at a time and the rest would be hidden, so the interface area looks tidier and more compact. This is implemented by define the same position for all the three panels but only one visible at once. The Fill button triggers the Panel section 2 (Figure 5-19) while hide the rest of the panels, and set the fields in this panel empty (String as empty and value as 0) so users can input his/her figures.

```
set(handles.uipanel_prediction,'Visible','off')
set(handles.uipanel_currentEnv,'Visible','on')
set(handles.uipanel_target,'Visible','off')
```

```
set(handles.uipanel_currentEnv,'Position',[52 0.8 80 22.3])

%%%%%%% initialize

set(handles.currentelec,'String','','Value',0)

set(handles.currentroad,'String','','Value',0)

set(handles.currentairlong,'String','','Value',0)

set(handles.currentairmedium,'String','','Value',0)

set(handles.currentairshort,'String','','Value',0)

set(handles.compareelec,'String','','Value',0)

set(handles.compareroad,'String','','Value',0)

set(handles.compareair,'String','','Value',0)
```

The Compare button compares the user input indicator values with the prediction from the neural networks, and gives the comparison result in the Edit Text boxes in Panel section 2. Furthermore, a data recording mechanism is implemented within the callback of this button. Every time a user uses this comparison button, the data he/she inputs will be written into a database, so more data can be collected every time somebody uses the system, and the neural nets and the system can be improved in the future.

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

```
conn = database('newdata', '', '');

colnames =
{'Employee','Teleworker','Computer','Office','Region','Teleconf','Electricity',...

   'Road','Airlong','Airmedium','Airshort'};

exportdata = {get(handles.Employee,'String'),get(handles.teleworker,'String'),...
```
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

```
get(handles.computers,'String'),get(handles.office,'String'),num2str(get(handles.Regio
n,'Value')),...
   get(handles.teleconf,'String'),get(handles.currentelec,'String'),...
   get(handles.currentroad,'String'),get(handles.currentairlong,'String'),...
   get(handles.currentairmedium,'String'),get(handles.currentairshort,'String')};
insert(conn, 'NewData', colnames, exportdata)
close(conn)
```

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

The Target button and the Panel 'Target and Improvements' (Figure 5-20) correspond to the expert system part. The drop down menus in the Targets Panel shows the range of each indicator to be chosen by the end users, once one is selected, the corresponding rule set will be presented in the Edit Text box. For example, if the end user chooses electricity target as between 15 and 20 units, the Edit Text box on the right will present the condition as "Employee Number - Teleworker Number) <= 20 and 12 < Computer Number <18 and 67.800003 < Office (sqrm) < 272 ')". This is implemented by the case switch statement.

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

```
switch strelec{valelec};
case '15 ~ 20'
set(handles.conditionselec,'String','(Employee Number - Teleworker Number) <= 20
and 12 < Computer Number <18 and 67.800003 < Office (sqrm) < 272 ');
```

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

Finally the Close button, the callback defined a pop up dialog asking the end user to confirm the close of the system.
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

```
user_response = modaldlg('Title','Confirm Close');
switch user_response
case {'No'}
   % take no action
case 'Yes'
delete(handles.figure1)
```

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

The implementation of the GUI was therefore completed.

## 5.7 Conclusions

This chapter has explained the Decision Support System built upon the Neural Network and Fuzzy Logic results from previous chapter, how this fit into the forward chaining of the expert decision support system. The implementation of the neural network based DSS was by development of three functions corresponding to the three modular networks.

Also the reasons that a backward chaining expert system was needed were presented. A comparison between Decision Support System and Expert System was made to explain why the expert system was a complement to the DSS and why this hybrid system was required. The process of building an expert system, the knowledge engineering, was conducted through the rule induction method and was implemented in the software Clementine. Results were presented in both decision tree and rule sets formats. The ES was validated by cross referencing to the original data, to see whether the data grouped together follow the rule for the group or not, and results were reasonable taken the confidence level of the rule into account.

The Graphical User Interface was built for convenience of end users of the EDSS, so understanding of the background technologies including Neural Network, DSS etc was not necessary to operate the system. The author described how the graphics was designed with the help of screenshots, and how the functions and callbacks were programmed to link the graphics with the background engine.

The generic DSS and ES concepts, components and mechanisms were explained. The development of the EDSS in this study, though might not follow the standard procedure and consists of standard elements, was surely for better integration of the DSS and ES, and customised to this research topic.

The background mechanism of the EDSS is the Neural Network application for analysing the relationship between e-business/ICT indicators and the environmental indicators, the knowledge discovery through rule induction. Although this background mechanism is the major scientific contribution of this study, the GUI is the final interface presenting the output of the study and the information-exchanging environment.

Construction of the forward chaining DSS and backward chaining ES is the first of its kind in this field, which has taken research in this topic to a brand new level, not only extended the dominating qualitative methods, but also expanded the quantitative methods to a new stage. The simulation capability has introduced a pilot scheme for this research ground.

Furthermore, with more end users use this EDSS, the recording mechanism has provided an adaptive feature to collect more data and the ability to improve the system all the way, and offering a promising start for further research in this area.

# Chapter 6

# EDSS Validation, Verification and Demonstration

## 6.1 Introduction

In Chapter 5, a hybrid expert system and decision support system was developed, based on three modular neural networks presented in Chapter 4 and knowledge discovery results. This chapter will further verify and validate this EDSS and demonstrate how an end user or a company/organisation can use and operate this expert decision support system for various purposes, such as estimation of $CO_2$ emission to stay within the carbon quota and comparisons of a company's current environmental performance with the average in the industry.

## 6.2 Validations and Verifications

In Computer Programming terminology, validation refers to "the process of controlling that data inserted into an application satisfies pre determined formats or complies with stated length and character requirements and other defined input criteria". Software validation is often confused with software verification, whose goal is to "assure that software fully satisfies all the expected requirements". The differences between them can be described as:

- Software verification asks the question, "Are we building the product right?"; that is, does the software conform to its specification.

- Software validation asks the question, "Are we building the right product?"; that is, is the software doing what the user really requires.

In this study, the validation and verification include: i) whether the predictions/outputs/results produced are right which means whether the error generated between the prediction and targets falls in an acceptable range. This is directly linked to whether the neural networks perform with acceptable error and whether the forward chaining DSS logic is right; ii) whether the conditions triggered to achieve certain defined targets are right. This is validation of the backward chaining logic; iii) whether the GUI constructed is reading and exchanging user inputs and calculated outputs from the right part of the engine. This is to verify that the final interface sends user inputs through to the right parts and presents back the right outputs.

### 6.2.1 Validation of Modular Neural Networks

Validation of three modular neural networks were conducted and explained in details in Chapter 4 as part of neural networks development, so it won't be repeated here. Just to summarise, the primary data - 100 sets of data - were used to train the network, while rest of the data - 20 sets of data chosen at random- were kept unseen to the network. The networks were firstly trained using the first group of data. Training errors (difference between the target and prediction of the network on the training data itself) were calculated. For the second group of data that was unseen to the networks, the errors were similarly calculated. This was then used to validate the network performance. The errors of the unseen data fell in an acceptable range and mostly were lower than the average training errors. Therefore the accuracy of the results and the network performances were reasonable, and generalisation ability of the neural network was validated.

### 6.2.2 Validation of Expert System Logics

Also the validation of the backward chaining Expert System logics was explained in Chapter 5 already, as part of expert system development itself. Just to summarise briefly, for the rule sets generated for the expert system, the rules were validated through cross referring back to the original datasets grouped together, to see whether they match the rules generated for that cluster or not. For most cases, all datasets follow the rules of the cluster to which the dataset belongs. Very few outliers do not follow the rules generated for the cluster these outliers belong to, however the instances were reasonable under the confidence level of the rule.

### 6.2.3 Verification of GUI

The previous chapter has explained the major implementation of the GUI and the connectivity. The following section will give an example to further test and verify the system. This part of verification is for the GUI, and connectivity between the graphical objects and the callbacks/functions in the codes.

For example, a company has 42 employees, 5 teleworkers, 51 computers, office space is 400 square meters, business operates internationally, and uses teleconferencing for an average of 45 hours a year. On the other side, the company consumes 50,000 kWh electricity per year, staff members travel 3000 km per week by road, has 3 times*persons long haul per year, and 8 times short haul. Now the company wants to see how good its environmental performance is in its industry sector. According to the calculation of neural networks built (reading and calculating directly using the neural network programmed codes), the prediction should be 42,150 kWh electricity, 4019.8 km road travel, 2 times long haul travel, 1 time medium haul, and 4 times short haul. The GUI should be able to produce same results. The procedure of validation is as follows.

Run/execute the GUI (which saved as ee.fig). Fill the employee number with 42, teleworker 5, computers 51, office space 400, region International, teleconferencing hour 45. Click the predict button. The results through this GUI interface are shown below, which are the same as the direct outputs calculated through neural network codes.

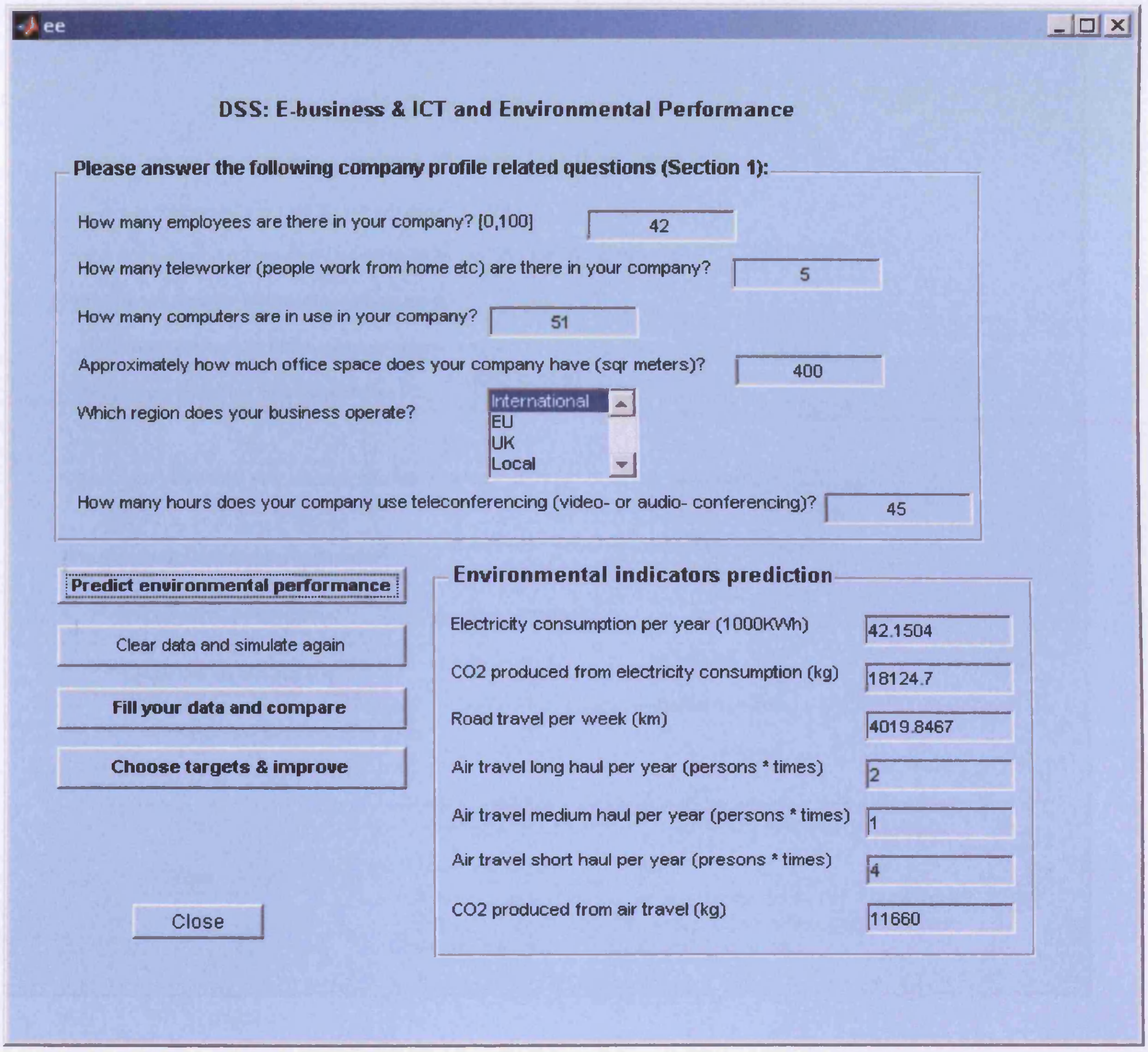



**Figure 6-1 Validation Prediction**

Also click the “Fill your data and compare” button. Fill in figures in the real case: for example, electricity as 50, road travel 3000, long haul 3, medium haul 0 and short haul 8, click the Compare button. Results of the comparison are shown as follow, which are also the right difference between the neural network predictions and the real cases.

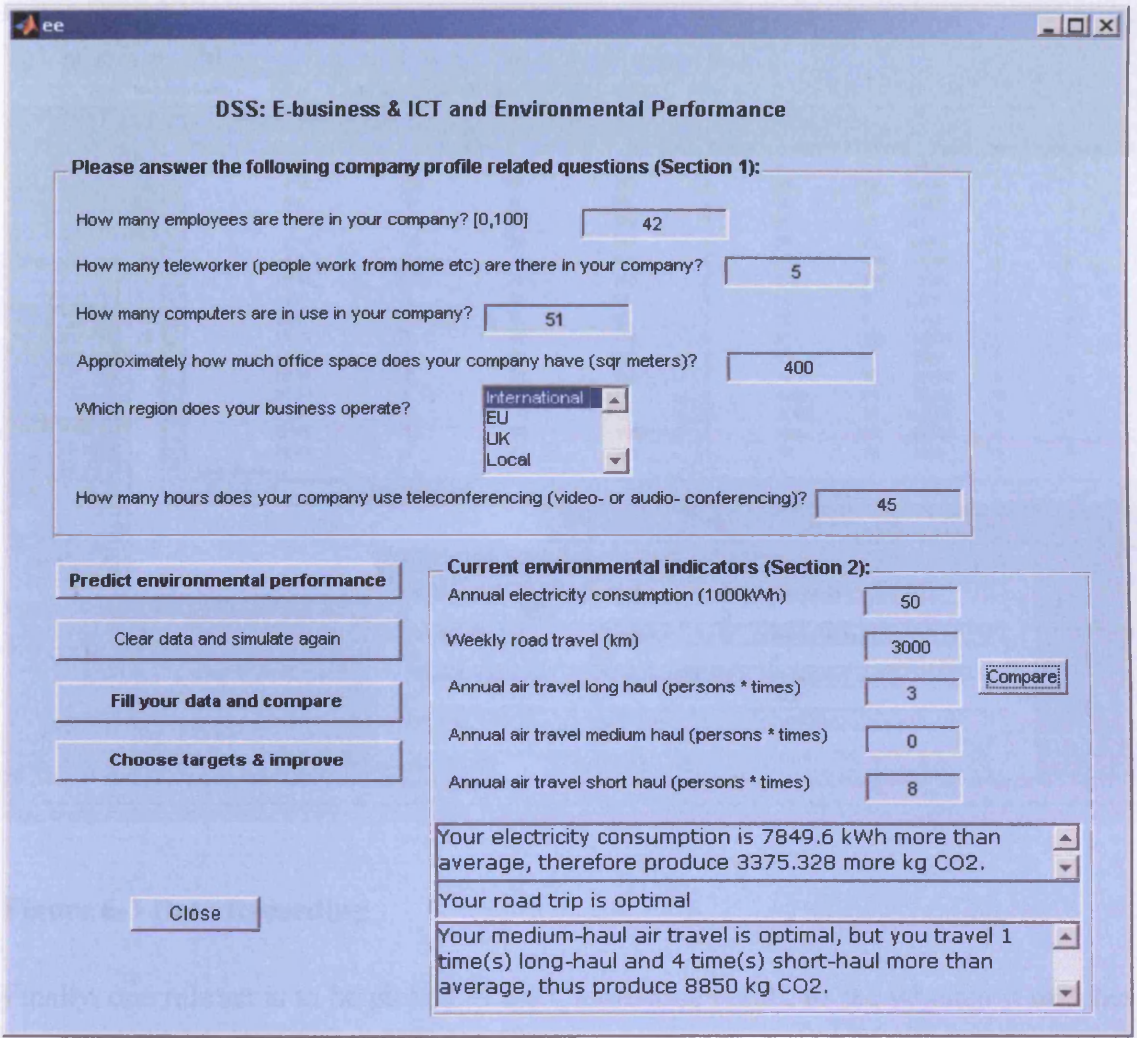



**Figure 6-2 Validation Compare**

Also as mentioned, there is a recording mechanism for collecting data that can help improve the system performance. If open the newdata.mdb database (Figure 6-3), the data just presented to the GUI can be seen in the database already (the last record), along with previous user inputs.

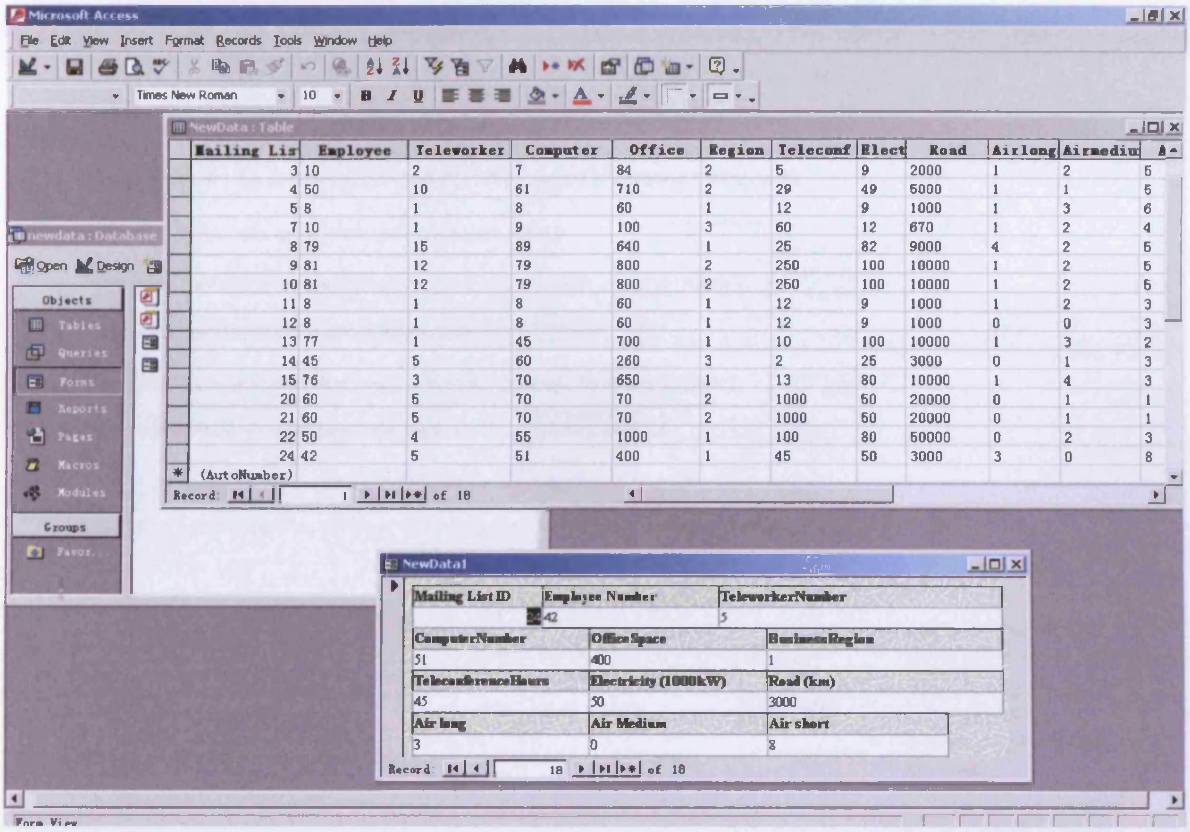



**Figure 6-3 Data Recording**

Finally, one rule set is to be picked in the Clementine result, to see whether it matches the system result. Suppose we choose road module, choose the road target as 2k to 3k, trace back to the original Excel file for road module, the cluster for [2k, 3k) is Cluster 2. In Clementine result, rule sets for Cluster 2 are:

Rules for 2 - contains 2 rule(s)
Rule 1 for **Cluster 2.0** (6, 0.75)
if TeleworkerNumber > 1
and em-tel > 24
and Office <= 232
then **Cluster 2.0**
Rule 2 for **Cluster 2.0** (12, 0.5)
if em-tel <= 24
and ComputerNumber <= 15
and Office > 61
and Office <= 232
then **Cluster 2.0**

On the other hand, results shown in the GUI the system (Figure 6-4), is the same as above.

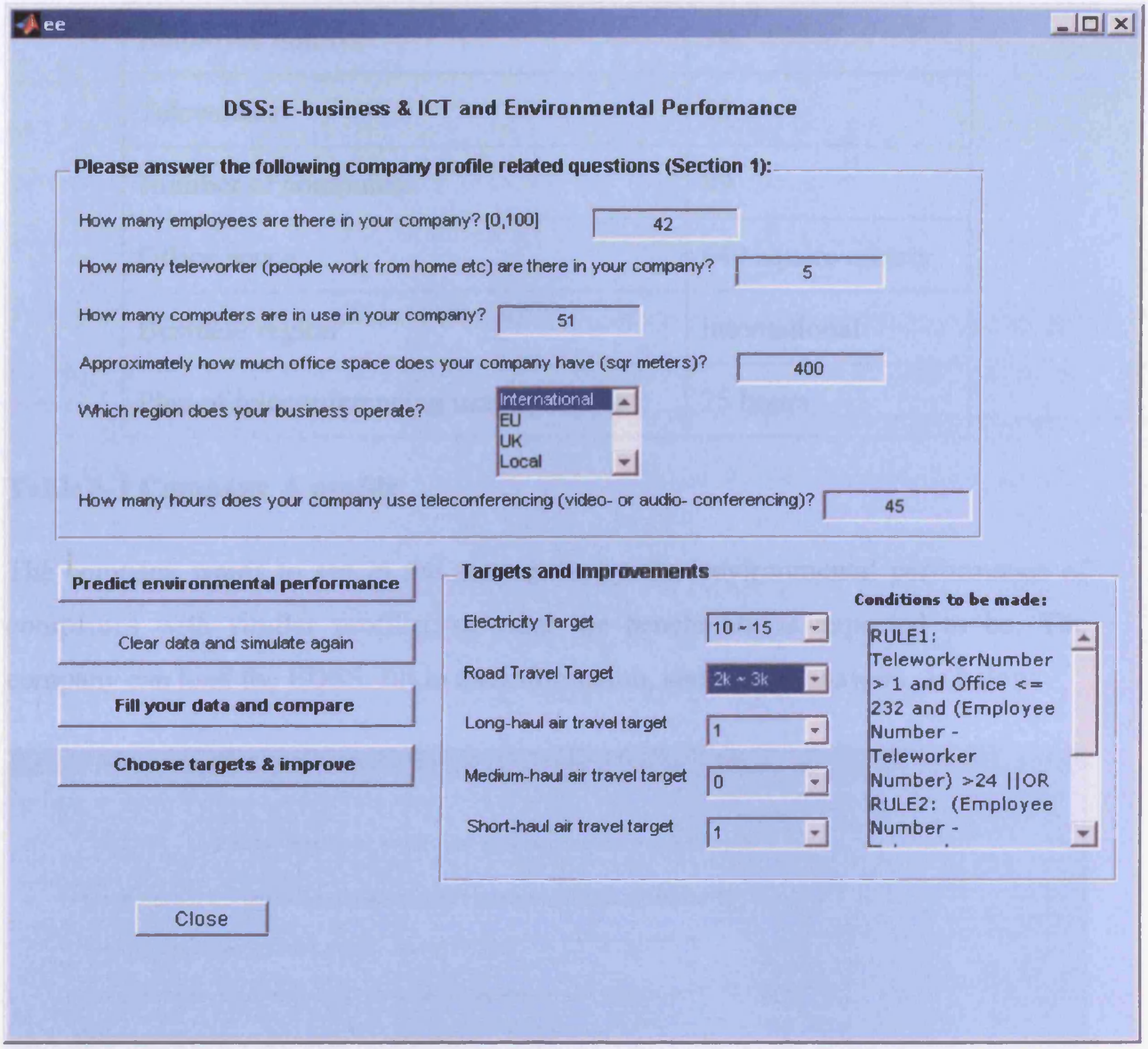



**Figure 6-4 Validation Targets and Conditions**

Therefore, the GUI and the function coding are validated to be successful.

## 6.3 Demonstration of the EDSS

This section will present how an end-user can use/operate this EDSS with different purposes.

### 6.3.1 Scenario 1 - Prediction

Company A just started its business. It has the following profile data:

| | |
|---|---|
| Employee number | 79 |
| Teleworker | 15 |
| Number of computers | 89 |
| Office space | 640 square meters |
| Business region | International |
| Plan of teleconferencing usage per year | 25 hours |

**Table 6-1 Company A profile**

The company wants to see in the same industry, the environmental performance of companies with similar profiles, or what the benchmark is expected to be. The company can load the EDSS, fill in the information, and get the answers, as below.

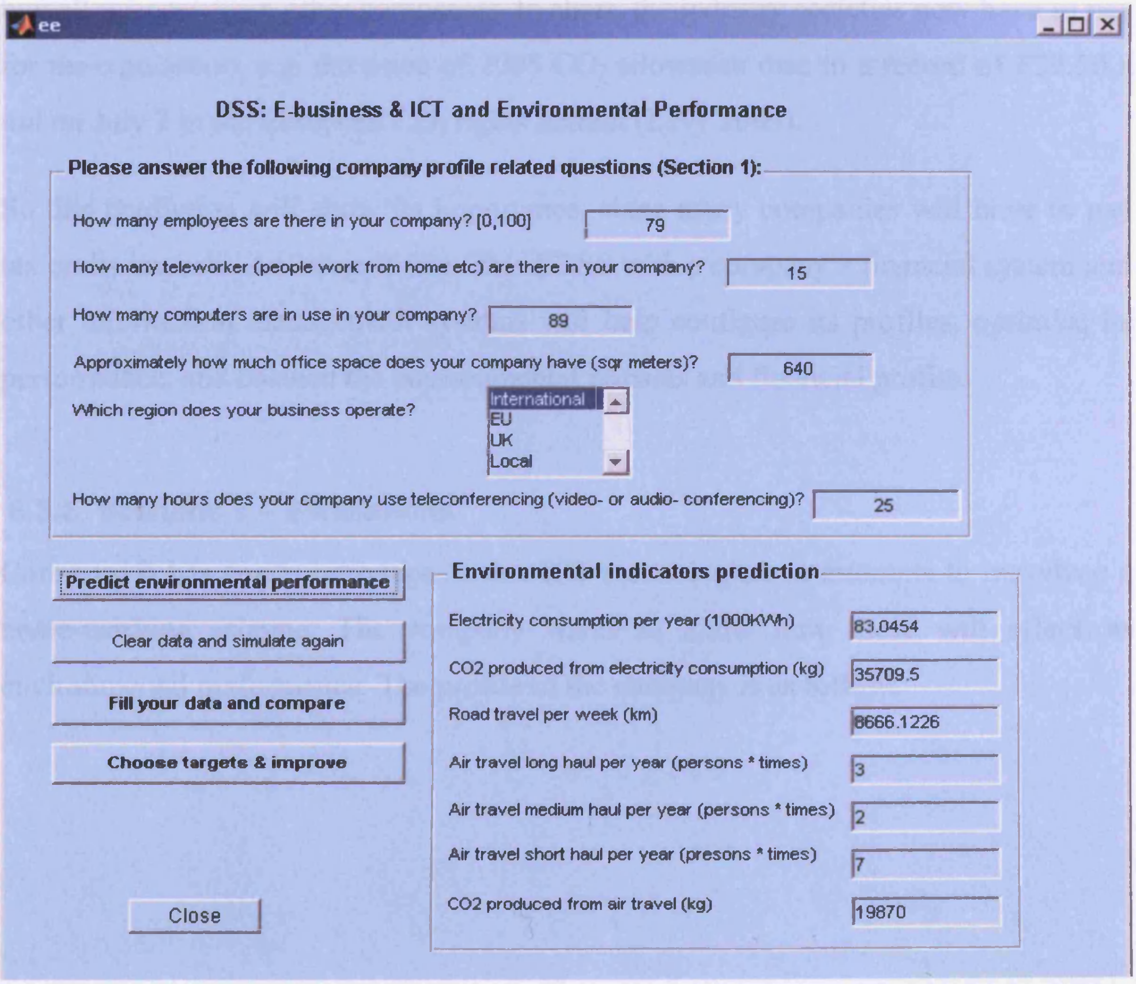



**Figure 6-5 Scenario 1**

These results tell the company, on average, how much electricity is consumed by a similar company in its industry sector, how much $CO_2$ is caused, how much road and air travel should be expected, and corresponding $CO_2$ from air travel. These data, representing the benchmark of the industry, are important for a start-up company to understand the industry's status and use for forecasting and pre-cautionary strategies.

Furthermore, provided with the rate of electricity from its supplier, the company can estimate how much it is expected to spend on electricity. This will also link to the company's financial profits. The reason is: driven by the Kyoto Protocol signed in 1997, Governments committed to reduce carbon dioxide ($CO_2$) emissions by 20% by 2008 – 2012, a green house gas the US Academy of Sciences says causes global warming. The British Government has introduced a tax on industrial energy as well as a cap and trade scheme. Companies are granted with annual allowance of $CO_2$ emission (the cap), whoever exceeds the cap would be fined; alternatively they can buy allowances from other companies. In short, the industry societies now have to pay for their pollution, e.g. the price of 2005 $CO_2$ allowance rose to a record of €29.50 a ton on July 7 in the European $CO_2$ rights market (Levy 2005).

So this prediction will show its importance, since many companies will have to pay tax on its impacts. An integration of this EDSS with a company's financial system and other information management systems will help configure its profiles, optimise its performance, and balance the environmental impacts and financial profits.

### 6.3.2 Scenario 2 – Simulations

Company B has newly introduced some ICT technologies, or attempts to introduce a home-working scheme. The company wants to know how these will affect its environmental performance. The profile of the company is as follow:

| | |
|---|---|
| Employee number | 35 |
| Plan of Teleworker number | 5 |
| Number of computers | 30 |
| Office space | 300 square meters |
| Business region | EU |
| Plan of teleconferencing usage per year | 45 hours |

**Table 6-2 Company B Profile**

First of all, the company can input "teleworkers number" as 5, the EDSS will provide estimation as follow (Figure 6-6). Next, it can keep all the other parameter values constant and change the "teleworker number" to 10 and see the changes (Figure 6-7).

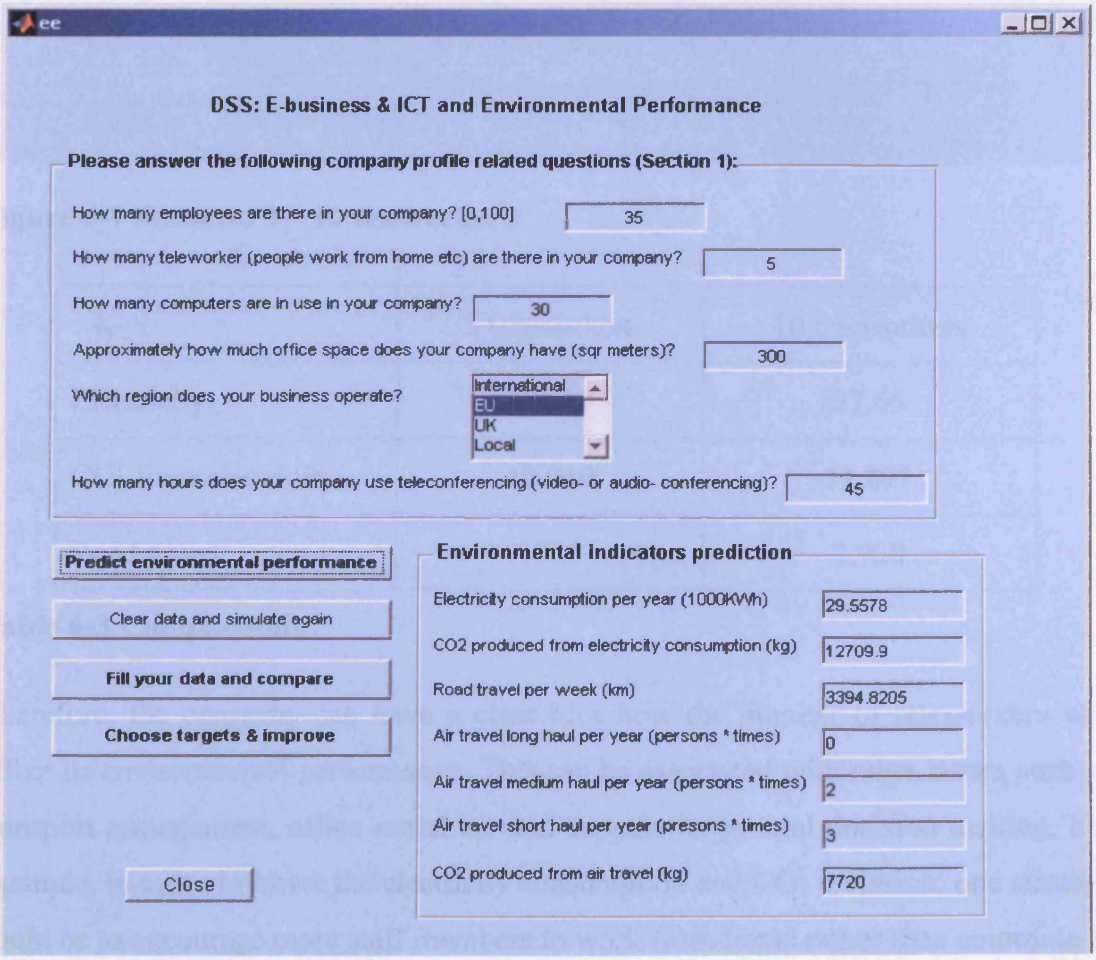



**Figure 6-6 Scenario 2 – 5 Teleworkers**

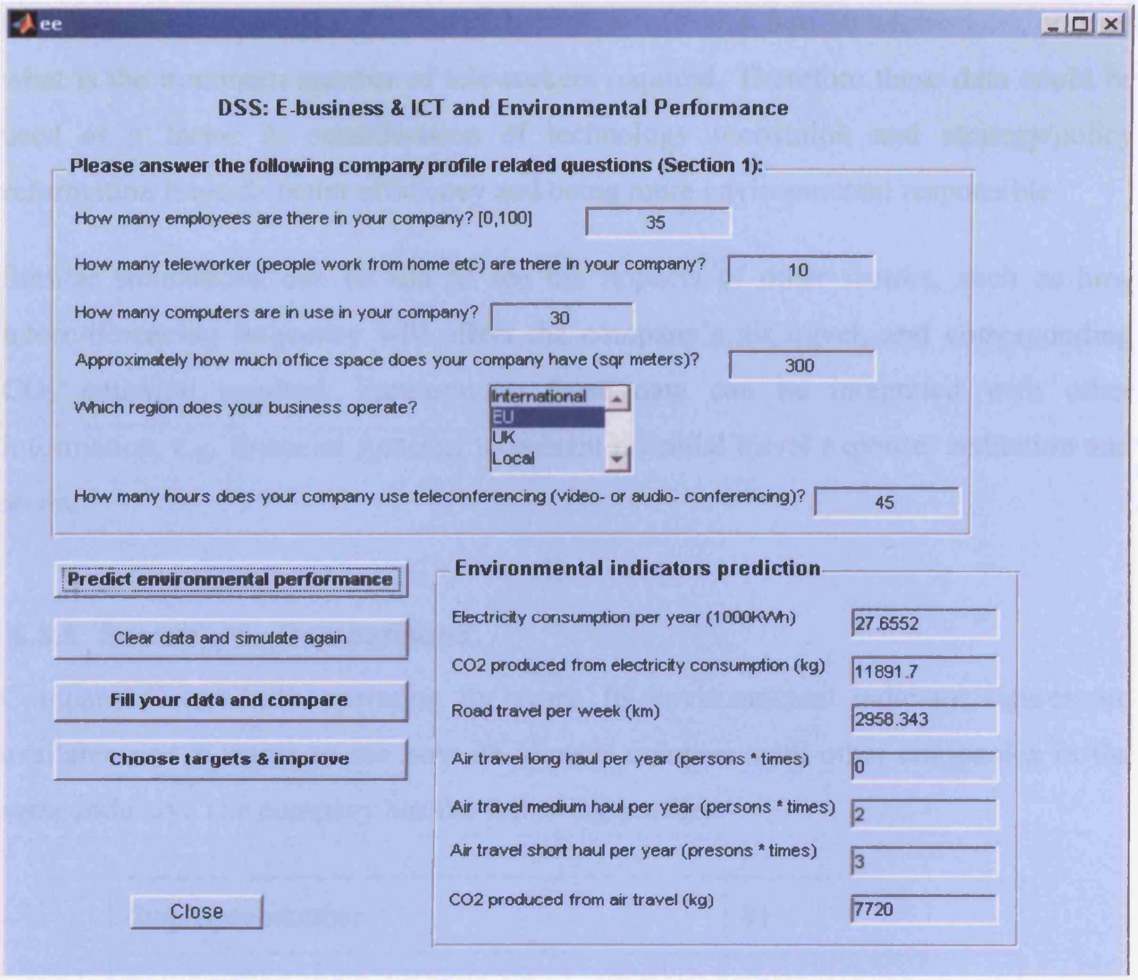



**Figure 6-7 Scenario 2 – 10 teleworkers**

| | 5 teleworkers | 10 teleworkers |
|---|---|---|
| Electricity | 29.56 | 27.65 |
| $CO_2$ from electricity | 12,710 | 11,891 |
| Road travel | 3395 | 2,958 |

**Table 6-3 Comparisons**

Therefore, the company can have a clear idea how the number of teleworkers will affect its environmental performance. This can be associated with other issues such as transport arrangement, office rental etc and support its general decision making. For example, in order to lower the electricity consumption and $CO_2$ emission, one strategy could be to encourage more staff members to work from home rather than commute to office; or for a company who intends to set the output electricity and $CO_2$ level to certain pre-defined targets, it can test the system with, i.e. 5 to 10 teleworkers, and see what is the minimum number of teleworkers required. Therefore these data could be used as a factor in consideration of technology innovation and strategy/policy reformation towards better efficiency and being more environmental responsible.

Similar simulations can be run to see the impacts of other factors, such as how teleconferencing frequency will affect the company's air travel, and corresponding $CO_2$ emission resulted. Furthermore, these data can be integrated with other information, e.g. financial systems, to present potential travel expenses reduction and so on.

### 6.3.3 Scenario 3 - Comparisons

Company C has been operating for years. Its environmental indicator figures are available and it wants to see how its impacts compare with other companies in the same industry. The company has the following profile.

| | |
|---|---|
| Employee number | 81 |
| Plan of Teleworker number | 12 |
| Number of computers | 79 |
| Office space | 800 square meters |
| Business region | EU |
| Plan of teleconferencing usage per year | 250 hours |
| Current electricity consumption (1000kWh) | 88 |
| Current road travel per week | 10,000 |
| Current long haul air travel | 1 |
| Current medium haul air travel | 3 |
| Current short haul air travel | 4 |

**Table 6-4 Company C Profile**

With the data input to the EDSS, the system will tell the company that:

- Your electricity consumption is 3549.6 kWh more than average, therefore producing 1526.328 more kg $CO_2$.
- You travel 693.1929 km per week more than average.
- Your short-haul air travel is optimal, but you travel 1 time(s) medium-haul and 1 time(s) long-haul more than average, thus producing 2810 kg more $CO_2$.

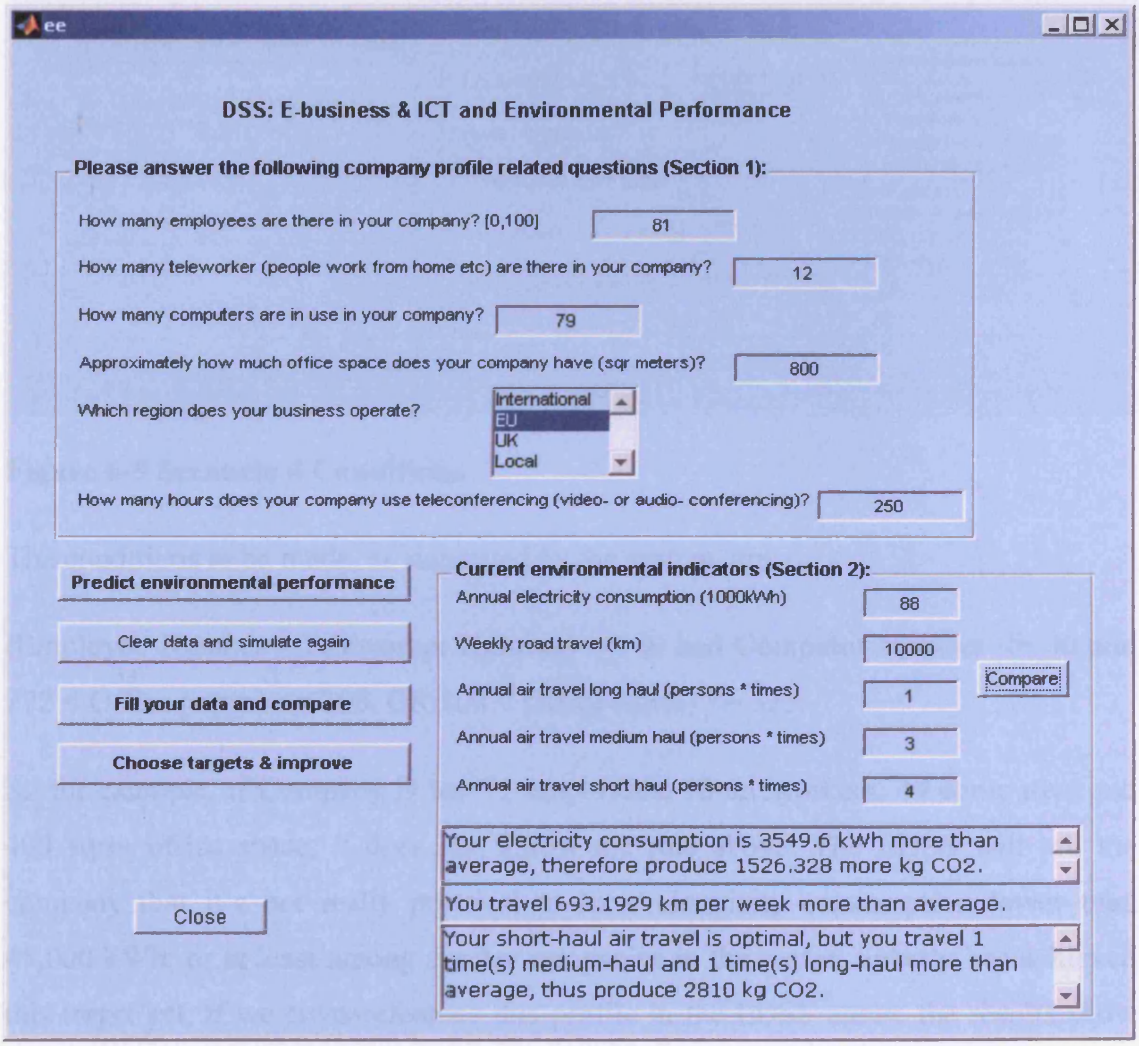



**Figure 6-8 Scenario 3 Comparison Results**

Therefore the company can understand in which area it has done well, which aspects need improvements and how much improvement can be expected to reach the benchmark. This information can be used to decide further action plans to achieve certain targets, and can be integrated to the company's strategic plan and assist further decision making.

### 6.3.4 Scenario 4 – Solutions

Company D wants to know what's in common between companies who consume 35 to 45 (unit: 1000kWh) electricity per year, so it can consider whether it is practical or not to fall into this category in the future. The EDSS will provide result as follow:

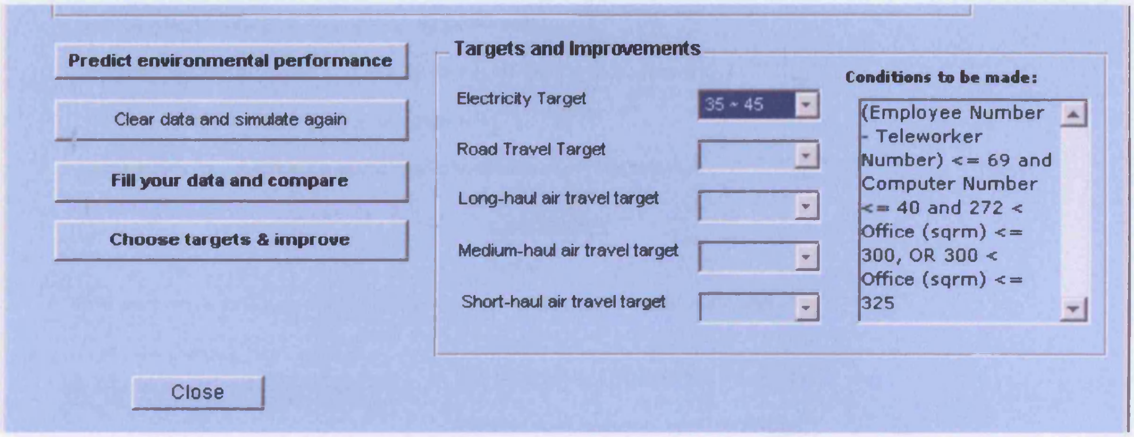



**Figure 6-9 Scenario 4 Conditions**

The conditions to be made, as suggested by the system, are:

(Employee Number - Teleworker Number) <= 69 and Computer Number <= 40 and 272 < Office (sqrm) <= 300, OR 300 < Office (sqrm) <= 325

So for example, if Company D has 75 employees, 10 teleworkers, 49 computers and 400 sqrm office space, it does not follow the rule above. The EDSS will tell the company that it's not really practical to have electricity consumption lower than 45,000 kWh; or at least among similar companies in the sector, nobody has achieved this target yet. If we cross-reference this profile in the EDSS again, the results show that the electricity consumption for a company like this will be 50,000 kWh (Figure 6-10). This also verifies the previous point that it will be an unprecedented case for a similar company to fall in the 35 ~ 45 category.

On the other hand, if Company E has 35 employees, 5 teleworkers, 35 computers and 200 sqrm office space, it does not follow this rule either. So this is not the category such a company should belong to either, but for a different reason to Company D. If we simulate this profile in the EDSS, results will show that the estimate of electricity consumption is 29,400 kWh. This tells us that companies with a similar profile to Company E should not produce 35 to 45 (1000 kWh), because it's too much.

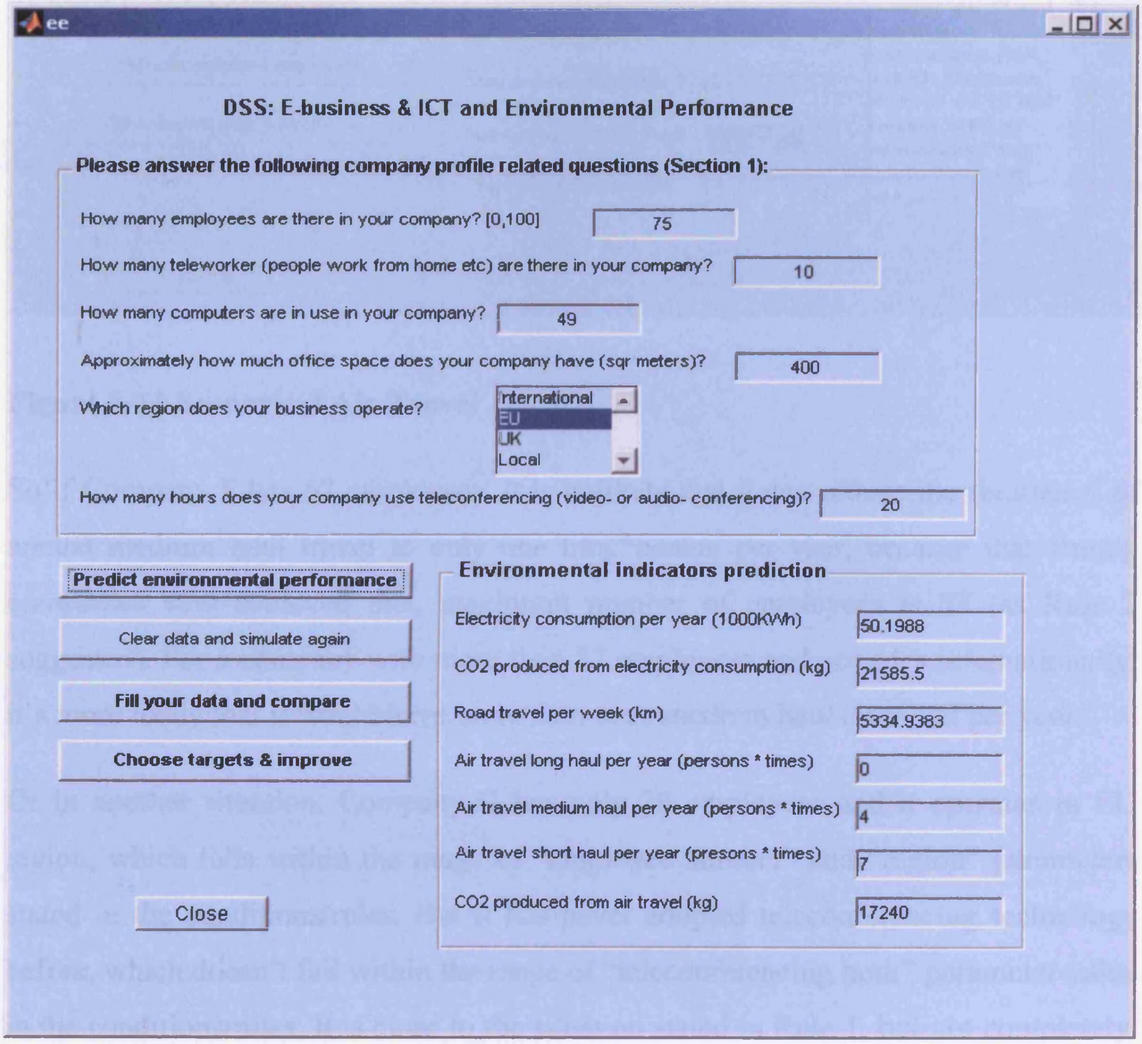



**Figure 6-10 Scenario 4 Average**

Or in another case, Company F wants to reduce its annual medium haul air travel to 1 time only; it will find the EDSS provides the following rules (Figure 6-11), which represent what's in common among companies who achieved one time medium haul travel only:

RULE1: 14 < Employee <= 37 and Region is EU and Teleconferencing > 0 hours

OR RULE2: Employee number <= 37 and region is international

OR RULE3: Employee Number <= 16 and regions is EU or international and teleconferencing usage is 0 hour

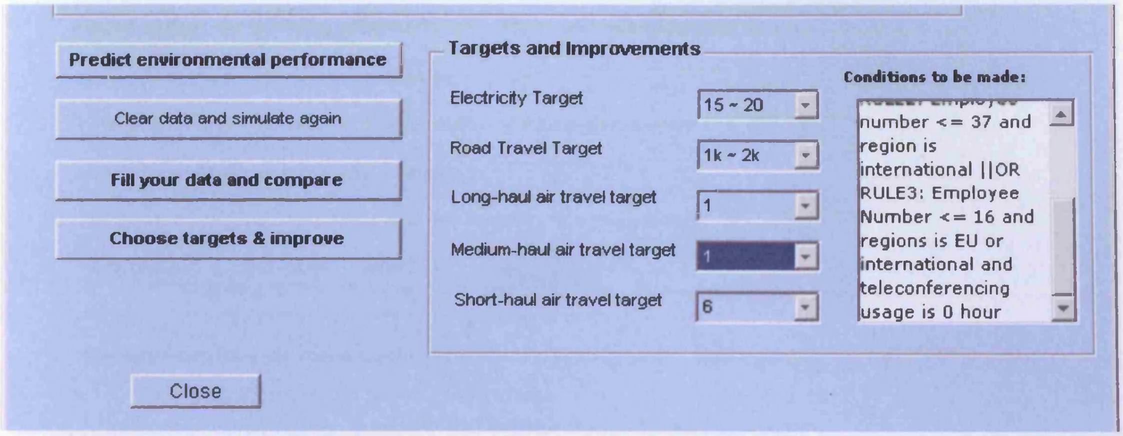



**Figure 6-11 Scenario 4 Air Travel**

So if Company F has 67 employees, it is unlikely that it can reduce the frequency of annual medium haul travel to only one time*person per year, because that among companies who achieved this, maximum number of employees is 37 (as Rule 2 suggested). For a company with more than 37 employees and operates internationally, it's more likely that it would have more than once medium haul air travel per year.

Or in another situation, Company G has only 28 employees and it operates in EU region, which falls within the range of "employee number" and "region" parameters stated in the conditions/rules. But it has never adopted teleconferencing technology before, which doesn't fall within the range of "teleconferencing hour" parameter value in the conditions/rules. It is close to the situation stated in Rule 1, but not completely. So if it wants to achieve this 1 time medium haul air travel target, probably it needs to start a strategy that applies teleconferencing into its business. As Figure 6-12 shows, it is likely that such a company will make 3 times medium haul travel a year without teleconferencing technology. This on the other hand, verified that in order to reduce this frequency to 1, applying teleconferencing technology would be needed.

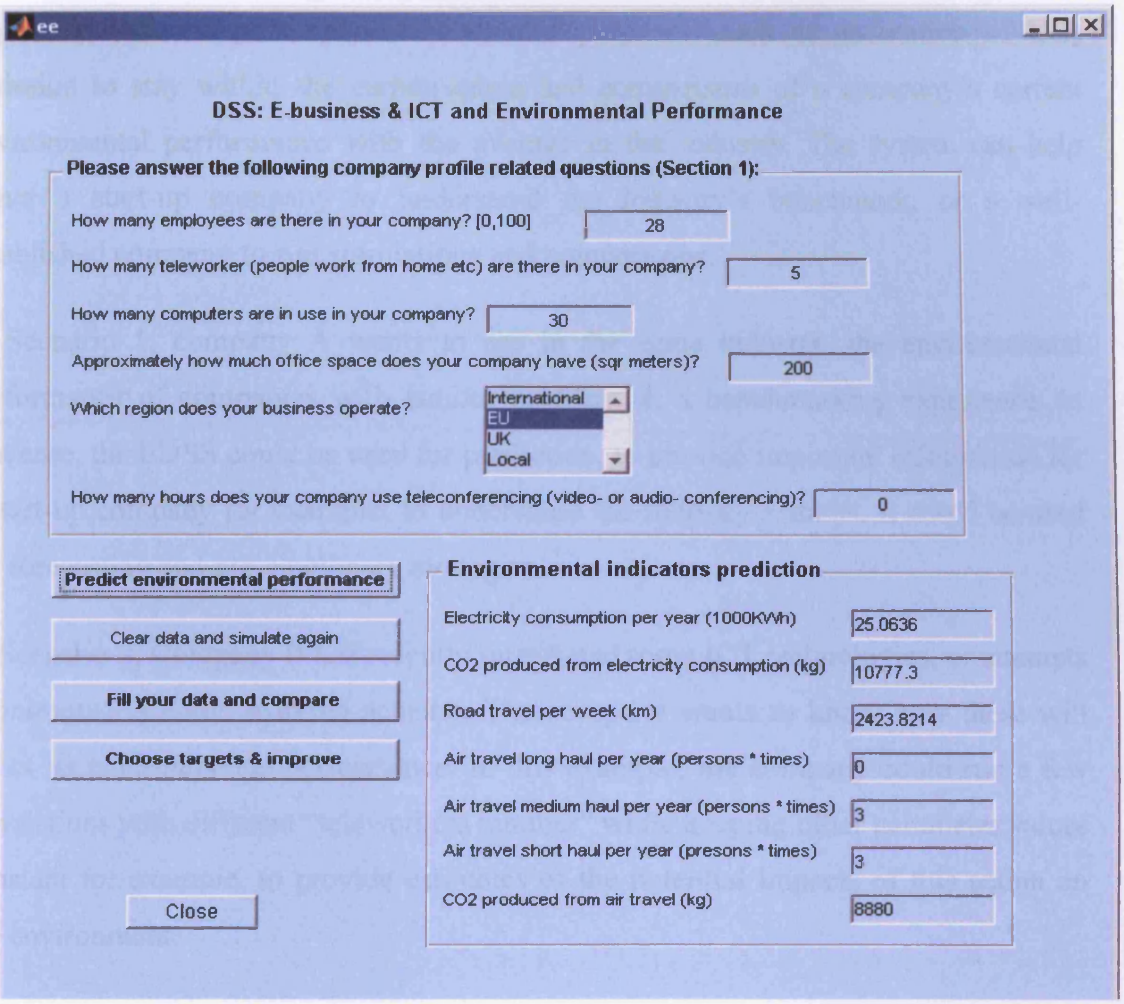



**Figure 6-12 Scenario 4, Company G**

## 6.4 Conclusions

This chapter has attempted to explain how the EDSS was validated and verified. Three parts of validation and verification were needed. These include, checks on the accuracy of the neural network engine, checks on the rules induced, and checks on the interface exchanges user inputs and advice properly with the background engine. The first two were conducted along with the implementation procedure and explained in detail in previous chapters. The third part was explained in detail in this chapter.

The latter part of the chapter presented four scenarios. Screenshots were provided to demonstrate how an end user or a company/organisation can use and operate this expert decision support system for various purposes, such as estimation of $CO_2$ emission to stay within the carbon quota and comparisons of a company's current environmental performance with the average in the industry. The system can help either a start-up company to understand the industry's benchmark, or a well-established company to run simulations and comparisons.

In Scenario 1, company A wants to see in the same industry, the environmental performance of companies with similar profiles i.e. a benchmarking experience. In that case, the EDSS could be used for prediction, to provide important information for a start-up company for example, to understand the industry's status. It could be used for forecasting and pre-cautionary strategies.

In Scenario 2, Company B has recently introduced some ICT technologies, or attempts to introduce a home-working scheme. The company wants to know how these will affect its environmental performance. In this example, the company could run a few simulations with different "teleworkers number" while keeping other parameter values constant for example, to provide estimates of the potential impacts of this action on the environment.

In Scenario 3, Company C has been operating for years. Its environmental indicator figures are available and it wants to see how its impacts compare with other companies in the same industry. In this scenario, the company uses the comparison function of the EDSS, to understand in which area it has done well, which aspects need improvement and how much improvement can be expected to reach the benchmark.

In Scenario 4, Company D wants to know what are the characteristics of companies who consume for example, 35 to 45 (unit: 1000kWh) electricity per year, so it can consider whether it is practical or not to fall into this category in the future. Or for another company, advice could be given regarding what conditions would be required to achieve certain targets.

All these data generated from the EDSS, are ideally expected to be integrated into a company's financial system and other information management systems. This important information could then be incorporated into a company's strategic plans, action plans and technological reformation.

# Chapter 7

# Conclusions and Future Work

## *7.1 Introduction*

So far salient conclusions specific to each chapter have been presented. The objective of this chapter is to now collate such specific conclusions, in order to present a holistic view of the concepts, importance and status of this research study. The following section will review the objective and tasks that were set out at initial stage of the study, summarise the background and status of the research field, and actions completed. Major contributions and suggestion of future work will also be presented.

## *7.2 Review of Objective and Tasks*

The principal aim of this thesis was to not only identify the major impacts of e-business and ICT on the environment, but also to demonstrate and develop an Expert Decision Support System, based on Fuzzy Logic and Neural Network technologies, at the meso level, to help a company/organisation analyse how e-business/ICT affects its environmental performance.

In order to achieve this, specific tasks as set out in chapter 1, were:

- An investigation of the state of the art identifying the relationship between e-business/ICT and the environment, and current methodologies, tools and models to analyse the relationship.

- The selection of an industry and the collection of data for relevant indicators representing the relationship.
- A qualitative and quantitative analysis of the data. Data mining and knowledge discovery will be employed to find trends and relationships in a data repository and hence build the knowledge base in this research field.
- The development of quantitative models to analyse the cause-effect relationship using Fuzzy Logic and Neural Networks, and target-conditions relationships using rule induction.
- The development of the Expert Decision Support System and the computer program/application encoding the models.

It is believed that each one of these tasks has been achieved and completed successfully. The following parts detail the main conclusions drawn from this work, main contributions and suggestions for further research.

## 7.3 Summary of the Background and Status of the Research Field

With less than half a century's development, e-business and the ICT technologies it relies on, have been growing rapidly and affecting profoundly the three pillars of sustainable development. With an even shorter history than the technology itself, the study of its impacts is still in its infancy. Recognition of the full impacts is rather limited but in progress. The e-revolution and the green revolution will proceed regardless of what sacrifice we might need to make and what 'side-effects' will come along, as they are improving our quality of life. It is better to learn to understand them so that pro-active and precautionary procedures can be taken if possible.

So far e-business and ICT have presented a positive picture in terms of economic and social impacts. However on the environmental side, the impacts are much more complicated and mixed. It is more difficult to define the boundary and isolate the factors of how ICT and e-business affect the environment, because more variables are involved, and above all, there are no approved standard indicators or quantitative approach to measure the effects.

Having reviewed literature in this field, the author realised that the research topic, while new and undefined, is promising and important. The two subjects involved (e-business/ICT and the environmental sustainability) are leading two revolutions in human history.

As expected from any new research field, awareness and understanding of this issue is limited. Past works have looked at the problem from limited specific dimensions, e.g. teleconferencing technologies and transport, or e-commerce and packaging waste. The various contributions to date have helped shape a promising and interesting research path. Researchers who have started looking into this issue have realised that the problem is not straightforward. This introduces new research challenges that are well worth studying further.

The importance of this research field has been gaining recognition day by day. Even though evidence so far has not presented uniform results, it has helped shape the research path. Current and traditional methodologies and tools to analyse the problem are limited.

It is acknowledged studies already exist, which develop quantitative indicators. The author concludes that the research community at the stage is trying to develop a perfect set of indicators for a specific field. Additionally it is acknowledged that a few models have been built to attempt to simulate certain specific parts of the relationship. However, the author concludes that these models have been developed based on a number of assumptions and have not yet been fully validated for major applications. In the author's opinion these are a good start but still need further consideration.

Overall, the author concludes that a more predictive and empirical model, which can be applied within a sector of society, should be more beneficial in the long term. Such an approach should help simulate potential impacts resulting from changes of indicator, so that positive effects can be promoted and negative ones alleviated.

## 7.4 Summary of Actions Completed

Though this study started trying to answer the question whether e-business/ICT affects the environment positively or negatively, it was soon realised that i) development of a knowledge base in this area and ii) demonstration of a new methodology to study the problem, were more important and urgent. This was because that i) data in this field is limited and ii) traditional methods and a systematic approach for impact studies have been found to be insufficient for this research topic due to its complicated nature.

Due to insufficient data and limited knowledge in this research field and in order to build a knowledge base for this research topic, a survey was conducted in the retail sector due to its increasing environmental responsibility and active position in utilising ICT technologies. However the data collected was rather insufficient therefore it was decided to seek further data from a second survey. Questionnaires were sent to members of the LRN project within the author's research centre, in the environmental industry. An online version of the questionnaire was designed in web page format and proved to be much more successful than the traditional mailshots. Databases were designed to store the data. The data was processed, and qualitative analysis and statistical analysis of the results were presented. However due to the relatively small data sample size, none specific conclusions were made as regard to whether e-business/ICT is "good" or "bad" for the environment, while the focus was on development and demonstration of a new methodology to study the problem domain and build a foundation for further studies.

Typical knowledge based techniques were investigated rather than traditional impact study techniques applied by other studies. The survey data were analysed further using various technologies, including Fuzzy Logic, Neural Networks, Decision Support System and Expert System techniques. These are well established technologies but have never been applied to this research area before.

Fuzzy logic was applied to convert linguistic terms in the survey data, such as increased/decreased, to numerical numbers that the system can understand and operate, as well as to reduce the data dimension. Less data was thus needed for neural network training. Also the survey data had missing values in some datasets, therefore data imputation techniques, data cleaning and normalisation were applied to pre-process and enhance the survey data before it was used in neural networks.

Neural networks, systems trained with and learned from survey data and able to forecast on unseen data, were used to model the survey data and analyse the relationship between e-business indicators and environmental performance indicators. However the data was still insufficient for one large sized network. So the idea was to break down the one overall network that includes all factors, inputs and outputs, to several modular networks, each dealing with certain specific aspect of the problem. Therefore a modular neural network approach was proposed and tested on some sample data. This approach was verified to be feasible and appropriate in this case. Three modular networks were chosen to represent the cases and demonstrate the idea. An electricity module, road travel module and air travel module were selected. The networks were constructed, trained and tested and optimal network structures were found to analyse the cause-effect relationship, i.e. which e-business/ICT parameter leads to which environmental results. This was used in the forward chaining decision support system. Overall the results have been satisfactory.

The target-conditions relationship, i.e. in order to achieve certain environmental performance, which e-business/ICT parameters need to be set, was analysed using Expert System technologies. The approach adopted included rule induction and other knowledge engineering techniques.

Finally the DSS and ES were integrated into a hybrid Expert Decision Support System, which was programmed in the Matlab environment. All the results, structure and knowledge discovered were integrated into a Graphical User Interface.

This EDSS was further validated and verified. Three stages of validation and verification are presented as follows: i) checks on the accuracy of the neural network engine, ii) checks on the rules induced and iii) checks that the interface exchanges user inputs and advice properly with the background engine.

A demonstration of how the system can be used and operated on various occasions for different purposes was presented, based on four application scenarios: predictions, simulations, comparisons and solutions. It was shown that an individual company can use this EDSS within a certain industry, for example, to simulate its environmental performance by adopting or limiting ICT technologies. Issues that can be explored include, for example, an estimation of $CO_2$ emission to stay within the carbon quota and/or comparisons of a company's current environmental performance with the average in the industry. The system can help either a start-up company to understand the industry's benchmark, or a well-established company to run simulations and comparisons. The data generated from the EDSS, are expected to be integrated into a company's financial system and other information management systems ideally. This important information could then be incorporated into a company's strategic plans, action plans and technological reformation.

## 7.5 Contributions

It is suggested that the major contributions of this study are as follow:

- Data in this field is rather scarce. Therefore it is critical for academics and governments to standardise indicators and methodologies and provide best practice for industry. The survey and data collection in this study are considered to be a contribution in this area.

- Data mining, data analysis and the knowledge discovery process to find trends and relationships in the data repository, was one of the core activities as well as one of the major contributions of this study. A knowledge base was built in this research field as a foundation for further studies.

- The research presents one of the pilot studies which try to not only build quantitative models but also to construct a Decision Support System to simulate this relationship in the real world. This work, it is claimed, extends current methodologies to a new level.

- The application of neural network technology in this research topic is a novel and innovative contribution of this study, being the first of its kind, not only introducing another methodology for this research ground, but also extending the traditional applications of neural network to a new challenge.

- The system developed is adaptive and capable of progressive data collection. Every time somebody uses the system, the recording mechanism, subject to confidentiality, will be able to collect data continuously, which can be used to re-train the neural networks and expand the knowledge base.

- Though the data used in the system was from one specific industry and the system should be used only within this industry, the infrastructure of the system is flexible and the procedures to build this EDSS are generic. Therefore data can be collected in another industry, used to re-train the neural networks and renew the system which can then be extended to other sectors.

- The research topic itself, as well as methodologies proposed to study the topic, covers various disciplines including business, IT, Artificial Intelligence, computer science and environmental science. It raised awareness of this interrelationship topic to a new level in the academics. The overall conclusions drawn indicate that the present work is an encouraging area for further research.

## 7.6 Future Work

An EDSS system almost, by definition, should be under constant development. Also it is recognised that a demonstration study in a pilot study attaches limitations to the work. The following suggestions are therefore made for further research.

- Exploring the full range of indicators. This PhD work has chosen certain indicators to represent only part of the relationship (air, electricity and road travel only). Therefore a generic methodology to solve this problem could be demonstrated. The real relationship covers wider aspects (for example, waste, water and many more) and a large range of impact factors. In order to understand the full picture, further investigations are needed.

- Collecting more data for the knowledge base development, to reduce some manual data processing and enhancement. However, this will need significant increase in the awareness and recognition of importance in this research problem itself.

- Once more data is available, an overall neural network instead of three modular neural networks is preferred, to avoid possible loss of information and contribution of insignificant factors and impacts.

- On the other hand if more data and knowledge were to be available, there could be an attribute in the EDSS called industry sector. So when the end user chose a certain industry sector, only the relevant questions and answers which constituted a modular EDSS that was specific to this sector would be triggered and presented.

- Incorporating the economic and social sustainability aspects in the system, to analyse the full range of relationships between e-business/ICT and the Sustainable Development.

- Transferring this stand-alone EDSS to an online application, so more users could access it. The knowledge base could thus be expanded more quickly.

# References


Abukhader, S. M. 2003. *The Environmental Implications of Electronic Commerce – The Assessment Approach Problem.* Licentiate dissertation, Lund University, Sweden.

Abukhader, S. M. and Jönson, G. 2003. The environmental implications of electronic commerce: a critical review and framework for future investigation. *Management of Environmental Quality: An International Journal* 14(4), pp. 460-476.

Abukhader, S. M. and Jönson, G. 2004. E-commerce and the environment: A gateway to the renewal of greening supply chains. *International Journal of Technology Management.*

Alakeson, V. et al. 2003. Social Responsibility in the Information Society *Digital Europe: ebusiness and sustainable development WP4*, p. 86.

Arnfalk, P. 1999. *Information Technology in Pollution Prevention.* Licentiate Dissertation, Lund University, Sweden.

Arnfalk, P. 2004. *The future impact of ICT on environmental sustainability: Evaluation and Recommendations (Fifth Interim Report)*
Lund: Institute for Prospective Technological Studies (IPTS).

Atlantic-Consulting 1998. *LCA Study of the Product Group Personal Computers in the EU Ecolabel Scheme*

Berkhout, F. and Hertin, J. 2004. De-materialising and re-materialising: digital technologies and the environment. *Futures* 36(8), pp. 903-920.

Brouwer, R. K. and Pedrycz, W. 2003. Training a feed-forward network with incomplete data due to missing input variables. *Applied Soft Computing* 3, pp. 23-36.

BT 2003. *E-business and the environment, Better World – Our Commitment to Society.*

Cabinet-Office 2005. Connecting the UK: the digital strategy. In: Cabinet-Office ed.

Cairns, S. 1999. *The home delivery of shopping: the environmental consequences.* London: ESRC Transport Studies Unit, University of London.

Caudill, R. et al. eds. 2001. *A life cycle environmental study of the impact of e-commerce on electronic products.* Proceedings of the Institute of Electrical and Electronics Engineering (IIIEE) International Symposium on Electronics and the Environment. USA:

Clark, R. et al. 2001. Retail Week's Top 500. London: emap retail.

Cohen, N. 1998. Greening the Internet: Ten Ways E-Commerce Could Affect the Environment and What We Can Do. *Environmental Quality Management* 9(1), pp. 1-16.

Condra, L. W. 1993. *Reliability Improvement With Design of Experiments.* New York: Marcel Dekker Inc. .

Cox, E. 1999. *The fuzzy systems handbook* 2nd ed. San Diego, USA: Academic Press, p. 716.

DEFRA 2000a. Environmental Reporting – Guidelines for Company Reporting on Greenhouse Gas Emissions. In: DEFRA ed. Crown.

DEFRA 2000b. Environmental Reporting Guidelines for Company Reporting on Water. In: DEFRA ed. Crown.

DEFRA 2000c. Environmental Reporting: Guidelines for Company Reporting on Waste. In: DEFRA ed. Crown.

DEFRA 2001. Environmental Reporting: General Guidelines. In: DEFRA ed. Crown.

Demuth, H. and Beale, M. 1992-2004. Neural Network Toolbox User's Guide Natick, MA: The MathWorks, Inc.

Digital-EU. 2006. *Digital Europe.* Available at: <URL: http://www.digital-eu.org> [Accessed: 22-08-2006 2006].

DigitalEurope. 2003. Ebusiness and sustainable regional development in Europe. *Digital Europe: ebusiness and sustainable development* [online]. Available at: <URL: http://www.digital-eu.org/>.

DTI 1994. *Best Practice Guidelines for Developing Neural Computing Applications* London: DTI.

DTI 1999. *e-commerce@its.best.uk, DTI Spectrum Survey* London: Cabinet Office.

Durieu, X. 2003. How Europe's retail sector helps promote sustainable production and consumption. *UNEP Industry and Environment* 26(1), pp. 7-9.

EC 2002. *e-business w@tch* Brussels: European Commission.

Eco-Label. 2006. *Eco-Label.* Available at: <URL: http://www.eco-label.com/default.htm> [Accessed: 18-09-2006].

Editorial 2003. Shopping for a better world: sustainability and retailing. *UNEP Industry and Environment* 26(1), pp. 3-6.

Ellis-Chadwick, F. et al. 2002. Signs of change? Alongitudinal study of Internet adoption in the UK retail sector. *Journal of Retailing and Consumer Services* 9, pp. 71-80.

Erdmann, L. and Behrendt, S. 2003. *The future impact of ICT on environmental sustainability: Script (Second Interim Report)*. Berlin: Institute for Prospective Technological Studies (IPTS).

Erdmann, L. and Würtenberger, F. 2003. *The future impact of ICT on environmental sustainability: Identification and global description of economic sectors (First Interim Report)*. Berlin: Institute for Prospective Technological Studies (IPTS).

European-Commission 2000. *Directive on waste electrical and electronic equipment and directive on the restriction of the use of certain hazardous substances in electrical and electronic equipment* Brussels: European Commission.

Feigenbaum, E. and McCorduck, P. 1983. *The Fifth Generation*. Reading, Mass. : Addison-Wesley.

Fichter, K. 2003. E-commerce - Sorting Out the Environmental Consequences. *Journal of Industrial Ecology* 6(2), pp. 25-41.

Fink, A. 1995. *The Survey Handbook*. The Survey Kit, SAGE Publications Inc.

Finlay, P. N. C., Mass., 1994. *Introducing decision support systems*. Oxford, UK Cambridge, Mass., NCC Blackwell; Blackwell Publishers.

Forum for the future. 2006. *Forum for the future*. Available at: <URL: http://www.forumforthefuture.org.uk/ > [Accessed: 22-08-2006 2006].

Frey, S. D. et al. 2000. Environmental Assessment of Electronic Products using LCA and Ecological Footprint. In: *Joint International Congress and Exhibition. Electronics goes green 2000*. Berlin,Germany.

Fulford, B. 2003. *Korea's Weird Wired World* Available at: <URL: Forbes.com > [Accessed: 05/04/2005].

Gay, R. H. et al. 2005. Modeling Paradigm for the Environmental Impacts of the Digital Economy. *Journal of Organizational Computing and Electronic Commerce* 15(1), pp. 61-82.

GeSI ed. 2002. *Industry as a partner for sustainable development: Information and Communication Technologies.* Paris, France: Global e-Sustainability Initiative and UNEP.

Giarratano, J. and Riley, G. 1998. *Expert Systems - Principles and Programming.* 3rd ed. Boston PWS Publishing Company, p. 597.

Goodman, J. and Alakeson, V. 2003. *The future impact of ICT on environmental sustainability: Scenarios (Third Interim Report).* London: Institute for Prospective Technological Studies (IPTS).

GRC. 2006. *Land Regeneration Network.* Available at: <URL: http://www.grc.cf.ac.uk/lrn/> [Accessed: 18-09-2006].

Green-Home. 2006. *Green Home.* Available at: <URL: http://www.greenhome.com/ > [Accessed: 22-08-2006 2006].

Han, J. and Kamber, M. 2001. *Data mining: concepts and techniques.* London, UK: Academic Press, p. 550.

Haykin, S. 1999. *Neural Networks - a comprehensive foundation.* 2nd ed. New Jersy: Prentice-Hall, Inc. .

Heinonen, S. et al. 2001. The ecological transparency of the information society. *Futures* 33(3-4), pp. 319-337.

Henry, J. J. et al. 1998. Neuro-fuzzy techniques for trafic control. *Control Engineering Practice* 6, pp. 755- 761.

Hermann, F. et al. 2001. Eco-Labelling and the Information Technology (IT) Industry. In: *the 2001 IEEE International Symposium on Electronics and the Environment.* Denver, Colorado.

Hesse, M. 2002. Shipping news: the implications of electronic commerce for logistics and freight transport. *Resources, Conservation and Recycling* 36(3), pp. 211-240.

Hilty, L. M. et al. 2006. The relevance of information and communication technologies for environmental sustainability - A prospective simulation study. *Environmental Modelling & Software* In Press, Corrected Proof.

Hilty, L. M. et al. 2004. *The future impact of ICT on environmental sustainability: Refinement and Quantification (Fourth Interim Report)*
St.Gallen: Institute for Prospective Technological Studies (IPTS).

Holt, B. and Benfer, R. A. 2000. Estimating missing data: an iterative regression approach. *Journal of Human Evolution* 39(289-296).

Hurst, E. 2001. *E-commerce and the Environment.* MSc Thesis, University of London.

IEEE. 2001. *The 2001 IEEE International Symposium on Electronics and the Environment.* Available at: <URL: http://ieeexplore.ieee.org/xpl/tocresult.jsp?isnumber=19985&isYear=2001# > [Accessed: 22-08-2006 2006].

IMRG. 2004. *UK ecommerce sales doubled in 2003.* Available at: <URL: http://www.imrg.org/IMRG/IMRGReports.nsf/(httpReportsByHotNews)/B0C6F22C509422B880256F5500392252 > [Accessed: 01/04/2005].

IMRG. 2006. *IMRG.* Available at: <URL: http://www.imrg.org> [Accessed: 22-08-2006 2006].

ISO14040 2006. Life Cycle Assessment - Principles and Guidelines.

ISO. 2006. *ISO 14001.* Available at: <URL: http://www.iso14000-iso14001-environmental-management.com/iso14001.htm> [Accessed: 18-09-2006].

IT and Environment Initiative. 2006. *IT and Environment Initiative.* Available at: <URL: http://www.it-environment.org/ > [Accessed: 22-08-2006 2006].

John, A. 2003. Information Technology and U.S. Energy Consumption: Energy Hog, Productivity Tool, or Both? *Journal of Industrial Econoloy* 6(2), pp. 13-24.

Jokinen, P. et al. 1998. The environment in an 'information society': A transition stage towards more sustainable development? *Futures* 30(6), pp. 485-498.

Jönson, G. et al. 2000. *E-commerce and Logistical Consequences (part 3).* Department of Design Science, Packaging Logistics at Lund Insititute of Technology.

Junninen, H. et al. 2004. Methods for imputation of missing values in air quality data sets. *Atmospheric Environment* 38, pp. 2895-2907.

Keskinen, A. et al. 2001. *Impacts of E-commerce on Transport and Environment* Joint Expert Group on Transport and the Environment to European Comission.

Kuhndt, M. et al. 2003. *Virtual Dematerialization and Factor X* Wuppertal Institute.

Levy, A. 2005. Making Pollution Pay. *Bloomberg Markets* September.

Lin, J. W. et al. 2003. A fuzzy neural network for assessing the risk of fraudulent financial reporting. *Managerial Auditing Journal* 18(8), pp. 657-665.

Luo, Y. et al. 2001. Network-based optimization and simulation of sustainable e-supply chain management. In: *Proceedings of the Institute of Electrical and Electronics Engineering (IIIEE) International Symposium on Electronics and the Environment.* USA.

Macauley, M. et al. 2003. Dealing with electronic waste: modeling the costs and environmental benefits of computer monitor disposal. *Journal of Environmental Management* 68(1), pp. 13-22.

Markeya, M. K. et al. In Press. Impact of missing data in evaluating artificial neural networks trained on complete data. *Computers in Biology and Medicine.*

MathWorks 1984-2001. Using Matlab.Natick: The MathWorks, Inc.

MathWorks. 1994-2006. *Neural Network Toolbox 5* Available at: <URL: http://www.mathworks.com/products/neuralnet/description1.html> [Accessed: 22-08 2006].

MathWorks 1995-2005. Fuzzy Logic Toolbox User's Guide Natick, MA: The MathWorks, Inc.

Matthews, H. S. et al. 2001. The net effect: environmental implications of e-commerce and logistics. In: *Proceedings of the 2001 Institute of Electrical and Electronics Engineering (IIEE) International Symposium on Electronics and the Environment.* Denver, Colorado.

Matthews, H. S. et al. 2002. Energy implications of online book retailing in the United States and Japan. *Environmental Impact Assessment Review* 22(5), pp. 493-507.

Miller, P. and Wilsdon, J. 2001. Digital Futures - An Agenda for a Sustainable Digital Economy. *Corporate Environmental Strategy* 8(3), pp. 275-280.

Mintzberg, H. et al. 1976. The structure of "unstructured" decision processes *Administrative Science Quaterly* 21, pp. 246-275.

Mokhtarian, P. L. 2003. Telecommunications and Travel: The Case for Complementarity. *Journal of Industrial Econoloy* 6(2), pp. 43-57.

Nelson, M. M. and Illingworth, W. T. 1991. *A Practical Guide to Neural Nets.* Reading, Massachusetts: Addison-Wesley Publishing Company, Inc. , p. 344.

Net-Impact-Study. 2002. The Net Impact Study overview. [online]. Available at: <URL: http://www.netimpactstudy.com/nis_2002.html > [Accessed: 15/05/2007].

New York Academy of Science. 2000. *Symposium at the New York Academy of Science on e-commerce and environment.* Available at: <URL: http://www.nyas.org/about/newsDetails.asp?newsID=102&year=2000> [Accessed: 22-08-2006 2006].

Norton, J. 2001. Winning in the race for e-business. *The Royal Academy of Engineering.*

OECD/ECMT. 2001. *Joint OECD/ECMT Seminar on the Impact of E-commerce on Transport* Available at: <URL: http://www.oecd.org/document/36/0,2340,en_2649_34351_2404580_1_1_1_1,00.html.> [Accessed: 22-08-2006 2006].

Park, J. and Roome, N. eds. 2002. *The Ecology of the New Economy: Sustainable Transformation of Global Information, Communications and Electronics Industries.* Greenleaf Publishing, p. 356.

Pe´rez, M. P. r. et al. 2004. The environmental impacts of teleworking: A model of urban analysis and a case study. *Management of Environmental Quality: An International Journal* 15(6), pp. 656-671.

Pesonen, E. et al. 1998. Treatment of missing data values in a neural network based decision support system for acute abdominal pain. *Artificial Intelligence in Medicine* 13, pp. 139-146.

Punakivi, M. and HolmstroÈ M, J. 2001. Environmental performance improvement potentials by food home delivery. In: *13th Nofoma Conference*. Reykjavik.

Quinlan, J. 1986. Induction of decison trees. *Machine Learning* 1, pp. 81-106.

Quinlan, J. R. 2000a. *Data Mining Tools See5 and C5.0.* Available at: <URL: http://www.rulequest.com/see5-info.html > [Accessed: 02-10-2006].

Quinlan, R. 2000b. *RuleQuest.* Available at: <URL: http://www.rulequest.com/> [Accessed: 02-10-2006].

Reijnders, L. and Hoogeveen, M. J. 2001. Energy effects associated with e-commerce: A case-study concerning online sales of personal computers in The Netherlands. *Journal of Environmental Management* 62, pp. 317 - 321.

Rejeski, D. 2003. E-Commerce, the Internet, and the Environment. *Journal of Industrial Econoloy* 6(2), pp. 1-3.

Romm, J. 2002. The Internet and the new energy economy. *Resources, Conservation and Recycling* 36(3), pp. 197-210.

Romm, J. et al. 1999. The Internet Economy and Global Warming: A Scenario of the Impact of E-commerce on Energy and the Environment Center for Energy and Climate Solutions (CECS), Global Environment & Technology Foundation (GETF).

Roome, N. and Park, J. 2000. Global Sustainability and Information Economy: Old Challenges, New Perspectives. *Greener Management International* 32, pp. 24-32.

Schultheis, R. A. and Summer, M. 1998. *Management Information System.* 4th ed. The McGraw-Hill Company, Inc. .

Scott Morton, M. S. 1971. *Management Decision Systems: Computer Based Support for Decision Making.* Division of Research Harvard University, Cambridge, MA., p. 216.

Siikavirta, H. et al. 2003. Effects of E-Commerce on Greenhouse Gas Emissions: A Case Study of Grocery Home Delivery in Finland. *Journal of Industrial Econoloy* 6(2), pp. 83-97.

SPSS 2003a. Clementine 8.1.Chicago.

SPSS 2003b. Clementine® 8.0 User's Guide.Chicago

SPSS 2005. Clementine® 10.0 - Specifications: Improve Results with High-Performance Data Mining. SPSS.

Sui, D. Z. and Rejeski, D. W. 2002. Environmental Impacts of the Emerging Digital Economy: The E-for-Environment E-Commerce? *Environmental Management* 29(2), pp. 155-163.

SustainIT. 2006. *SustainIT.* Available at: <URL: http://www.sustainit.org/ > [Accessed: 22-08-2006 2006].

SVTC. 2006. *Silicon Valley Toxics Coalition* Available at: <URL: http://www.svtc.org/> [Accessed: 22-08-2006 2006].

Taiariol, F. et al. 2001. Environmental impact of a telecommunication service. In: *The 2001 IEEE International Symposium on Electronics and the Environment.* Denver, Colorado.

Tarassenko, L. 1998. *A Guide to Neural Computing Applications.* New York: Arnold and John Wiley & Sons Inc.

Toffel, M. W. and Horvath, A. 2004. Environmental Implications of Wireless Technologies: News Delivery and Business Meetings. *Environmental Science & Technology* 38(11), pp. 2961-2970.

Tongia, R. et al. 2004. *Information and Communications Technology for Sustainable Development: Defining a Global Research Agenda.* Bangalore, India: Carnegie Mellon University, Pittersburgh & Indian Institute of Science, Banglore.

Turban, E. 1988. *Decision Support and Expert Systems - Manageral Perspectives* New York Macmillan Publishing Company, p. 697.

Turban, E. 1995. *Decision Support and Expert Systems : Management Support Systems* Englewood Cliffs, N.J.: Prentice Hall.

University-of-Abertay-Dundee. 2006. *Video Conferencing.* Available at: <URL: http://iserv.tay.ac.uk/av/vidconf/intro.html> [Accessed: 18-09-2006].

Wager, P. A. et al. 2005. Smart labels in municipal solid waste - a case for the Precautionary Principle? *Environmental Impact Assessment Review* 25, pp. 567-586.

Wikipedia. 2006. *Expert System.* Available at: <URL: http://en.wikipedia.org/wiki/Expert_system> [Accessed: 26-09-2006].

Wilsdon, J. ed. 2001. *Digital Futures: Living in a Dot-Com World.* London: Earthscan, p. 228.

Yi, L. and Thomas, H. R. 2006. E-business and Sustainable Development. *International Journal of Environment and Sustainable Development* 5(3), pp. 262-274.

Zadeh, L. A. 1965. Fuzzy Sets. *Information Control* 8, pp. 338-353.

Zhong, M. et al. 2004. Estimation of missing traffic counts using factor, genetic, neural, and regression techniques. *Transportation Research Part* C(12), pp. 139-166.

# Appendix A: Questionnaire 1

Miss Lan Yi
Geoenvironmental Research Centre
(ENGIN 1)
Cardiff University
PO Box 925
Cardiff
CF24 0YF

Date: 28th October 2003

Dear Sir / Madam,

**Re: Survey of the Relationship Between E-commerce & the Environment**

I write to invite you to participate in a survey being conducted by the Geoenvironmental Research Centre (GRC), School of Engineering, Cardiff University on the interrelationship between e-commerce and the environment.

In October 2000 the Prime Minister challenged the top 350 UK companies to produce environmental reports by the end of 2001. It is likely that this will eventually become a compulsory requirement for all companies in the near future.

With your help in this survey, this project will assist us in creating a Generic Decision Support System that could help companies (such as your company) produce environment reports whilst helping them to realize a). what opportunities/threats are imposed by "the environment" on their business; and b). how their e-business activities impact the environment.

The enclosed questionnaire has been specifically designed for ease of use and should only take approximately 10 to 15 minutes to complete. Should you require any assistance or input in completing the survey please do not hesitate to contact Miss Lan Yi at 029 20874645 or YiL@cf.ac.uk .

All information will be treated confidentially and any identifying information will be removed in order to protect commercial confidentiality.

It would be appreciated if you could please return the completed questionnaire with the stamped envelope provided at your earliest convenience to Miss Lan Yi.

Thank you very much for your cooperation.

Yours faithfully

*Lan Yi*
**Geoenvironmental Research Centre**

## Survey of "Relation Between E-Commerce & the Environment"

*Geoenvironmental Research Centre, School of Engineering, Cardiff University*

GRC CARDIFF UNIVERSITY

1. **Your information**

| | |
|---|---|
| **a) Name** | **b) Title** Prof / Dr / Mr / Mrs / Ms / Miss |
| **c) Position** | **d) Tel** |
| **e) Email Address** | |

2. **Company / Organisation Profile**

| | |
|---|---|
| **a) Company / Organisation name** | |
| **b) Address (inc. Postcode)** | |
| **c) Number of Employees** | **d) Year when company was founded** |
| **e) Industry Sector** | **f) Number of outlets** |
| **g) Affiliation: Subsidiary of** | |
| **h) Principal products/services** | |
| **i) Principal Market** | |

**Questionnaire** Instruction: Please tick the boxes that match most your answers, or give a short answer (mostly are figures) in the blank field.

3. **Environment policies, legislations and systems**

**a) How do you rank your company's attitude towards environment protection? 1: no relation with our company, ..., 5: very important to our company**

1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

**b) How important are the following environment related factors to your company? Please give a number from 1 to 5 to each. ( 1: no relation, ..., 5: very important)**

| | | | |
|---|---|---|---|
| Disposal of dangerous waste | ☐ | Reduction of packaging, non-hazardous waste | ☐ |
| Management of paper waste | ☐ | Minimization of energy (electricity, gas, water) costs | ☐ |
| Minimization of atmosphere waste | ☐ | Land contamination prevention | ☐ |
| Noise control | ☐ | Others, please specify ________ | ☐ |

**c) Does your company intend to join the EU Eco-label scheme?** [1]

[1] EU Eco-label: A voluntary scheme to enable European consumers to easily identify officially proved green products across the European Union.

Yes ☐ No ☐ Already implemented ☐ Do not Know ☐

**d) Does your company intend to obtain ISO 14001 Certificate?** [2]

Yes ☐ No ☐ Already implemented ☐ Do not Know ☐

**e) Do you have any Environment Management System in place?**

Yes ☐ No ☐ Do not Know ☐

**If no, why?**

Simply do not care about these issues ☐ No one has asked us to put such systems in place ☐
No in-house expert to address these issues ☐ Not enough time to think about these issues ☐
Not enough money to address these issues ☐ Do not have any significant environmental impact ☐

**f) Does your company audit environmental performance?**

Yes, every ________ year(s) ☐ No ☐ Do not Know ☐

**g) What are the main reasons for your pro-environment actions?**

Environment pressure from the government ☐ Procedure centered on quality: applied for ISO 14001 ☐
Pressure from customers ☐ Decreasing costs ☐
From head office ☐ Others, please specify ________________ ☐

**h) What do you do with waste electronic and electrical equipments (WEEE)?**

Recycle ☐ Disposal ☐ Charity donation ☐ Give to staff for free ☐ Do not know ☐

**i) Do you prefer likeminded environment-friendly suppliers?**

Yes ☐ No ☐ Do not Know ☐

**Do you still prefer these suppliers even if they cost you more?**

Yes ☐ No ☐ Do not Know ☐

**j) In October 2000 the Prime Minister challenged the top 350 UK companies to produce environmental reports[3] by the end of 2001. Do you think this should be applied to your company?**

Yes ☐ No ☐ Already implemented ☐ Do not Know ☐

**k) Does your company prefer to use recycled materials, such as recycled paper?**

Yes ☐ No ☐ Already implemented ☐ Do not Know ☐

## 4. Resources and outcome of environment protection actions

[2] ISO 14001: A voluntary international standard that defines the system elements an organisation needs to effectively manage its impact on the environment. Key components include: a companywide environmental policy; a systematic approach to planning, implementation and operations management; regular management reviews; checking and corrective action processes

[3] Environmental Reporting involves the publication in a company's Annual Report and/or self-standing reports of general environmental policy statements, usually including details of environmental performance such as greenhouse gas emissions, waste, water use and other relevant impacts giving quantified data and improvement targets.

**a) What are the profitable pro-environment actions?**

Reduction of raw materials ☐ Reduction of energy consumption ☐
Efficient energy conversion ☐ Others, please specify ________ ☐

**b) What resources did your company devote to the environment?**

Manager devotes part of his/her time ☐ Budget is part of the sale ☐

**c) How much waste (approximately) does your company send for disposal per year?** ______ **tonnes**

**d) (In order to estimate your company's Green House Gas and $CO_2$ contribution)**

**How much (approximately) do you spend on electricity per year?** ______

**How much (approximately) do you spend on gas per year?** ______

**How much (approximately) do you spend on water per year?** ______

**What is the quantity of fossil fuels (approximately) used per year for transport?** ______

**e) How have the above figures changed since you started conducting business via the electronic network?** **1: decreased a lot, .., 3: not much change ,...., 5: increased a lot**

| | | | | | |
|---|---|---|---|---|---|
| Electricity | 1 ☐ | 2 ☐ | 3 ☐ | 4 ☐ | 5 ☐ |
| Gas | 1 ☐ | 2 ☐ | 3 ☐ | 4 ☐ | 5 ☐ |
| Water | 1 ☐ | 2 ☐ | 3 ☐ | 4 ☐ | 5 ☐ |
| Fossil fuels | 1 ☐ | 2 ☐ | 3 ☐ | 4 ☐ | 5 ☐ |

**f) How much does your company spend on environment related actions per year?**

0 ☐ 0 ~ £ 500 ☐ £ 500 ~£ 1,000 ☐ £ 1,000 ~ £ 5,000 ☐ £ 5,000 ~ £ 10,000 ☐
£ 10,000~£ 50,000 ☐ more than £ 50,000 ☐ ☐ ☐

**g) What (how much) is your annual turnover in year 2002? (Optional) £** ________

5. **Human Resource Management and environment**

**a) Does your company have any specific person responsible for pro-environment issues?**

Yes ☐ No ☐ Do not Know ☐

**b) If yes, does she/he hold another position?**

Yes ☐ No ☐ Do not Know ☐

**c) If yes, what else position does she/he hold?** ________

**d) What do you do to train staff to become more environmentally-aware?**

Bulletin board in cafeteria ☐ Regular training ☐ Suggestion box ☐
Environmental Committee ☐ Employee evaluation takes environment into account ☐

6 **E-commerce information**

**a) What percentage of SALES is via e-commerce?** ________ %

**b) What percentage of PURCHASES is via e-commerce?** ________ %

**c) Since when has your company started using electronic network to conduct business?** ________

**d) What electronic media are used in e-commerce communication?**

E-mail ☐ Internet ☐ Telephone ☐ Fax ☐ EDI (Electronic Data Interchange) ☐

**e) Does your company have its own website?**

Yes, since year ________ ☐ No ☐ Do not Know ☐

**f) If yes, what do you use it for?**

Providing information ☐ Communicate with customers ☐ Communicate with suppliers ☐
Online sales ☐ Online purchases ☐ Support market campaigns ☐

**g) If yes, what is the website address (URL)?** ________

7. **Interaction of e-commerce and environment**

**a) How has paper consumption changed in your company since it started using the electronic network to conduct business? 1: decreased a lot, ..., 3: not much change, ..., 5: increased a lot**

1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

**b) Do you have any intention of replacing some travelling with teleconferencing?**

Yes ☐ No ☐ Already implemented ☐ Do not Know ☐

**c) Do you have any intention of replacing commuting with telecommuting?** e.g. staff work at home, which saves office space, time and transportation

Yes ☐ No ☐ Already implemented ☐ Do not Know ☐

**d) Do you intend to use renewable resources?** e.g. use small, silent electric vans for distribution, with less congestion, less noise and less pollution, which are powered by renewable resources

Yes ☐ No ☐ Already implemented ☐ Do not Know ☐

**e) How does doing business electronically contribute to a pro-environment attitude in your company? 1: very negative,** e.g. e-business increases energy consumption and contributes to global warming, **..., 3: not much change, ..., 5: very positive,** e.g. E-business saves energy, paper use, transportation fee ...

1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

**f) How does a pro-environment requirement contribute to your business? 1: very negative,** e.g. it gives more pressure and hassle to our company **..., 3: not much change, .., 5: very positive,** e.g. it provides opportunity to strengthen company reputation, brand image etc

1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

# Appendix B: Questionnaire 2

Shawn Green - IT Manager
Hayes Engineering Services Ltd
Brindley Road,
Cardiff,
CF11 8TL,

Dear

**LRN: Survey of the Relationship Between E-Business & the Environment**

I write to seek your support as a member of the Land Regeneration Network to undertake a short questionnaire on the interrelationship between e-business and the environment.

Your help in this survey will assist in developing a National Assembly funded project for a *Generic Decision Support System* that could help companies, such as yours, monitor and report environmental performance more effectively, reducing future costs.

The data from this survey will be used to train the decision support system's neural networks to help guide the use of e-business and resources in the environmental sector.

The enclosed questionnaire has been specifically designed for ease of use and should only take approximately 5 to 10 minutes to complete. Should you require any assistance or input in completing the survey please do not hesitate to contact either of my colleagues - Dr Martyn Jones (029 2087 6814 /JonesMJ7@cf.ac.uk); or Miss Lan Yi (029 2087 4000 ext 77401 / YiL@cf.ac.uk).

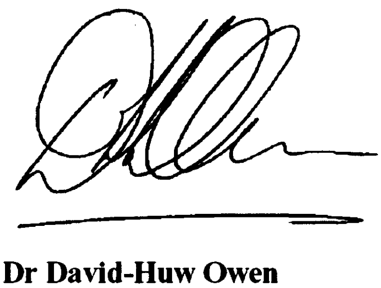
**PLEASE NOTE: an online submission system is also available at www.grc.cf.ac.uk/e-survey/**

All information will be treated confidentially and any identifying information will be removed in order to protect commercial confidentiality.

I would very much appreciate it if you could please complete the questionnaire at your earliest convenience and before 15th December 2004.

Thank you very much for your cooperation.

Yours sincerely



**Dr David-Huw Owen**
**Development Manager**
**Geoenvironmental Research Centre**

Dr David-Huw Owen
Geoenvironmental Research Centre
(ENGIN 1), Cardiff University,
PO BOX 925, Cardiff CF24 0YF

Telephone +044 (0) 29 2087 6697
Fax +044 (0) 29 2087 4004
E-mail OwenDH@cf.ac.uk
Web Site www.grc.cf.ac.uk

# Survey of "Relation Between E-Business & the Environment"

*Geoenvironmental Research Centre, School of Engineering, Cardiff University*

Please complete this form, giving as much detail as possible and return to the address provided. To complete the form online please go to http://www.grc.cf.ac.uk/e-survey/

*Please note: All information will be treated confidentially and any identifying information will be removed in order to protect commercial confidentiality.*

| | Company / Organisation Profile |
|---|---|
| 1.1 | **Company / Organisation name** |
| 1.2 | **Industry Sector** *(tick one box only)*: Engineering Consultant ☐ Environmental Service Provider ☐ Manufacturer ☐ Local Authority ☐ Waste Management ☐ Engineering Contractor ☐ Others ☐ |
| 1.3 | **Is you company** *(tick one box only)*: products supplier ☐ OR service supplier ☐. |
| 1.4 | **Is your business** *(tick one box only)*: international ☐ UK ☐ Wales ☐ local ☐ |
| 1.5 | **Gross annual turnover last year?** £ |
| 1.6 | **Number of Employees** |
| 1.7 | **Year when company was founded** |
| 1.8 | **Contact name** *(optional)* **Email** *(optional)* **Tel** *(optional)* |

| | E-business related information |
|---|---|
| 2.1 | **How many computers (inc. back-office infrastructure) does the company have?** ............ |
| 2.2 | **On average how many hours per day are the computers left on in your office?** ............ Hours per day |
| 2.3 | **Does your company have its own website?** Yes ☐ No ☐ *(please skip to question 2.4)*<br>**If yes, which year was your first website set up?** ............<br>**If yes, what use is made of the website?** *(tick as many boxes as apply)*<br>Providing information ☐ Communicating with customers ☐ Communicating with suppliers ☐<br>Online sales ☐ Online purchases ☐ Support market campaigns ☐<br>**If yes, what is the website address (URL)?** ............ |
| 2.4 | **How many tele-workers and/or home workers are there in your company?** ............ |
| 2.5 | **Does your company use tele-conferencing?** Yes ☐ No ☐ Don't know ☐<br>**If you do, how many TOTAL HOURS is it used per year?** ............ persons × hours |
| 2.6 | **What electronic media are used in your communication?** *(tick as many boxes as apply)*<br>E-mail ☐ Internet ☐ Telephone ☐ Fax ☐ EDI (Electronic Data Interchange) ☐ |
| 2.7 | **How have the following factors changed in your company since it started using Internet and related technologies to conduct business?** *(tick one box for each)* |

*1: decreased a lot 2: decreased slightly 3: not much change 4: increased slightly 5: increased a lot*

| | | | | | |
|---|---|---|---|---|---|
| Paper consumption | 1 ☐ | 2 ☐ | 3 ☐ | 4 ☐ | 5 ☐ |
| Electricity consumption | 1 ☐ | 2 ☐ | 3 ☐ | 4 ☐ | 5 ☐ |
| Gas consumption | 1 ☐ | 2 ☐ | 3 ☐ | 4 ☐ | 5 ☐ |
| Supply chain life cycle | 1 ☐ | 2 ☐ | 3 ☐ | 4 ☐ | 5 ☐ |
| Inventories | 1 ☐ | 2 ☐ | 3 ☐ | 4 ☐ | 5 ☐ |
| Land use | 1 ☐ | 2 ☐ | 3 ☐ | 4 ☐ | 5 ☐ |
| Employee number | 1 ☐ | 2 ☐ | 3 ☐ | 4 ☐ | 5 ☐ |
| Choice of suppliers | 1 ☐ | 2 ☐ | 3 ☐ | 4 ☐ | 5 ☐ |
| Customer number | 1 ☐ | 2 ☐ | 3 ☐ | 4 ☐ | 5 ☐ |
| Business transaction cost | 1 ☐ | 2 ☐ | 3 ☐ | 4 ☐ | 5 ☐ |
| Packaging | 1 ☐ | 2 ☐ | 3 ☐ | 4 ☐ | 5 ☐ |
| Business travels | 1 ☐ | 2 ☐ | 3 ☐ | 4 ☐ | 5 ☐ |

| | |
|---|---|
| 2.8 | **Please estimate the percentage of SALES via e-commerce in your company:** .........% or ☐ **None** |
| 2.9 | **Please estimate the percentage of PURCHASES via e-commerce in your company:** .........% or ☐ **None** |
| 2.10 | **How does '*doing business electronically*' contribute to a pro-environment attitude in your company?**<br>*KEY: 1: very negative, e.g. e-business increases energy consumption and contributes to global warming, ...,*<br>*3: not much change, ..., 5: very positive, e.g. E-business saves energy, paper use, transportation fee ...*<br>1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ |

| | Environment related information |
|---|---|
| 3.1 | **Has your company obtained ISO 14001 Certificate?** Yes ☐ No ☐ Don't know ☐ |
| 3.2 | **Has your company implemented an Environment Management System?** Yes ☐ No ☐ Don't know ☐<br>**If no, why?** *(tick as many boxes as apply)*<br>Simply do not care about these issues ☐ No one has asked us to put such systems in place ☐<br>No in-house expert to address these issues ☐ Not enough time to think about these issues ☐<br>Not enough money to address these issues ☐ Do not have any significant environmental impact ☐ |
| 3.3 | **Does your company produce Environment Report or Corporate Social Responsibility (CSR) report?**<br>Yes ☐ No ☐ Don't know ☐ |
| 3.4 | **Does your company audit environmental performance?**<br>Yes, every ..........year(s) ☐ No ☐ Don't know ☐ |
| 3.5 | **Please estimate what is the area of company office space:** ..........$m^2$ or ..........sqft |
| 3.6 | **If applicable, please estimate what is the area of company warehouse space:** ..........$m^2$ or ..........sqft |
| 3.7 | **Please estimate how much waste your company sends for disposal per year:** ..........tonnes |
| 3.8 | **Please estimate how much waste your company recycles per year?** ..........tonnes |
| 3.9 | **Please estimate how much electricity is consumed by your company per year:** ..........KWh or ..........£ |
| 3.10 | **Please estimate how much gas is consumed by your company per year:** ..........KWh or ..........£ |
| 3.11 | **Please estimate the distances travelled in total by your company for business purposes:**<br>**a. by road *per <u>week</u>*?** ..........km or ..........miles<br>**b. by train *per <u>week</u>*?** ..........km or ..........miles<br>**c. by air *per <u>year</u>*?** ..........km or ..........miles |

Please note: All information will be treated confidentially and any identifying information will be removed in order to protect commercial confidentiality.

Should you require any assistance or input in completing the survey please do not hesitate to contact

- Dr Martyn Jones (029 2087 6814 /JonesMJ7@cf.ac.uk); *or*
- Miss Lan Yi (029 2087 4000 ext 77401 / YiL@cf.ac.uk)

***Please return the completed questionnaire at your earliest convenience:***

***Online submission at: www.grc.cf.ac.uk/e-survey/***

***Or by fax to:*** *029 20870045*

***Or by post to:*** *Dr Martyn Jones, Geo-environmental Research Centre, Cardiff University, PO Box 925, Cardiff, CF24 0YF.*

**Thank you very much for your cooperation!**

# Appendix C: Clementine Rule Induction

**Electricity module:**

Rules for 0 - contains 1 rule(s)

Rule 1 for 0.0 (4, 0.833)

if ComputerNumber <= 6 and Office <= 67.800003 then 0.0

Rules for 1 - contains 1 rule(s)

Rule 1 for 1.0 (9, 0.909)

if ComputerNumber > 6 and Office <= 67.800003 then 1.0

Rules for 2 - contains 1 rule(s)

Rule 1 for 2.0 (8, 0.7)

if em-tel <= 20 and ComputerNumber <= 12 and Office > 67.800003 then 2.0

Rules for 3 - contains 1 rule(s)

Rule 1 for 3.0 (10, 0.75)

if em-tel <= 20 and ComputerNumber > 12 and Office > 67.800003 then 3.0

Rules for 4 - contains 1 rule(s)

Rule 1 for 4.0 (5, 0.857)

if em-tel > 20 and em-tel <= 35 and ComputerNumber <= 18 then 4.0

Rules for 5 - contains 1 rule(s)

Rule 1 for 5.0 (11, 0.615)

if em-tel > 20 and em-tel <= 35 and ComputerNumber > 18 and Office <= 272 then 5.0

Rules for 6 - contains 1 rule(s)

Rule 1 for 6.0 (6, 0.75)

if em-tel > 35 and Office <= 272 then 6.0

Rules for 7 - contains 2 rule(s)

Rule 1 for 7.0 (3, 0.8)

if Office > 300 and Office <= 325 then 7.0

Rule 2 for 7.0 (9, 0.727)

if em-tel <= 69 and ComputerNumber <= 40 and Office > 272 then 7.0

Rules for 8 - contains 2 rule(s)

Rule 1 for 8.0 (11, 0.923)

if EmployeeNumber <= 94 and ComputerNumber > 40 and ComputerNumber <= 55 and Office > 325 then 8.0

Rule 2 for 8.0 (47, 0.367)

if Office > 272 then 8.0

Rules for 9 - contains 2 rule(s)

Rule 1 for 9.0 (10, 0.75)

if EmployeeNumber <= 94 and ComputerNumber > 55 then 9.0

Rule 2 for 9.0 (11, 0.692)

if EmployeeNumber <= 94 and em-tel > 69 then 9.0

Rules for 10 - contains 1 rule(s)

Rule 1 for 10.0 (8, 0.8)

if EmployeeNumber > 94 then 10.0

Default: 8

## Road travel module:

Rules for 0 - contains 2 rule(s)

Rule 1 for 0.0 (4, 0.833)

if em-tel <= 4 then 0.0

Rule 2 for 0.0 (2, 0.75)

if em-tel <= 15 and ComputerNumber > 15 then 0.0

Rules for 1 - contains 4 rule(s)

Rule 1 for 1.0 (7, 0.889)

if em-tel > 4 and Office <= 61 then 1.0

Rule 2 for 1.0 (6, 0.75)

if EmployeeNumber > 11 and ComputerNumber <= 15 and Office <= 125 then 1.0

Rule 3 for 1.0 (2, 0.75)

if em-tel > 4 and ComputerNumber <= 6 then 1.0

Rule 4 for 1.0 (9, 0.727)

if em-tel > 15 and em-tel <= 24 and ComputerNumber > 15 then 1.0

Rules for 2 - contains 2 rule(s)

Rule 1 for 2.0 (6, 0.75)

if TeleworkerNumber > 1 and em-tel > 24 and Office <= 232 then 2.0

Rule 2 for 2.0 (12, 0.5)

if em-tel <= 24 and ComputerNumber <= 15 and Office > 61 and Office <= 232 then 2.0

Rules for 3 - contains 3 rule(s)

Rule 1 for 3.0 (3, 0.8)

if em-tel <= 32 and Office > 232 then 3.0

Rule 2 for 3.0 (8, 0.8)

if TeleworkerNumber > 5 and Office > 232 and Office <= 370 then 3.0

Rule 3 for 3.0 (7, 0.556)

if TeleworkerNumber <= 1 and em-tel > 24 and Office <= 232 then 3.0

Rules for 4 - contains 1 rule(s)

Rule 1 for 4.0 (11, 0.692)

if em-tel > 32 and em-tel <= 68 and ComputerNumber <= 45 and Office > 232 then 4.0

Rules for 5 - contains 2 rule(s)

Rule 1 for 5.0 (4, 0.667)

if TeleworkerNumber > 5 and em-tel <= 68 and ComputerNumber > 45 and Office > 370 then 5.0

Rule 2 for 5.0 (5, 0.571)

if TeleworkerNumber <= 5 and em-tel <= 60 and ComputerNumber > 45 and Office > 232 then 5.0

Rules for 6 - contains 1 rule(s)

Rule 1 for 6.0 (3, 0.6)

if TeleworkerNumber <= 5 and em-tel > 60 and em-tel <= 68 then 6.0

Rules for 7 - contains 1 rule(s)

Rule 1 for 7.0 (6, 0.625)

if em-tel > 68 and Office <= 540 then 7.0

Rules for 8 - contains 1 rule(s)

Rule 1 for 8.0 (20, 0.409)

if em-tel > 68 then 8.0

Rules for 9 - contains 2 rule(s)

Rule 1 for 9.0 (4, 0.833)

if TeleworkerNumber <= 9 and em-tel > 85 then 9.0

Rule 2 for 9.0 (2, 0.75)

if EmployeeNumber <= 70 and em-tel > 68 then 9.0

Default: 1

## Air Travel module

**Short-haul travel**

Rules for 0 - contains 1 rule(s)

Rule 1 for 0.0 (5, 0.857)

if Region <= 1 then 0.0

Rules for 1 - contains 2 rule(s)

Rule 1 for 1.0 (3, 0.8)

if Employee <= 11 and Region > 1 and TeleScale <= 0 then 1.0

Rule 2 for 1.0 (26, 0.75)

if Employee <= 37 and TeleScale > 0 then 1.0

Rules for 2 - contains 1 rule(s)

Rule 1 for 2.0 (5, 0.857)

if Employee > 11 and Employee <= 20 and Region > 1 and TeleScale <= 0 then 2.0

Rules for 3 - contains 2 rule(s)

Rule 1 for 3.0 (9, 0.727)

if Employee > 37 and Employee <= 64 and TeleScale > 1 then 3.0

Rule 2 for 3.0 (8, 0.7)

if Employee > 20 and Employee <= 37 and Region > 1 and TeleScale <= 0 then 3.0

Rules for 4 - contains 1 rule(s)

Rule 1 for 4.0 (53, 0.2)

if Employee > 37 then 4.0

Rules for 5 - contains 1 rule(s)

Rule 1 for 5.0 (14, 0.312)

if Employee > 37 and Employee <= 64 and TeleScale <= 1 then 5.0

Rules for 6 - contains 2 rule(s)

Rule 1 for 6.0 (4, 0.667)

if Employee > 64 and TeleScale > 3 and TeleScale <= 4 then 6.0

Rule 2 for 6.0 (6, 0.5)

if Employee > 37 and Employee <= 64 and Region <= 2 then 6.0

Rules for 7 - contains 1 rule(s)

Rule 1 for 7.0 (12, 0.429)

if Employee > 64 and Region > 2 and TeleScale <= 3 then 7.0

Rules for 8 - contains 1 rule(s)

Rule 1 for 8.0 (2, 0.75)

if Employee > 50 and Employee <= 64 and Region <= 2 and TeleScale <= 0 then 8.0

Default: 1

**Medium-haul travel**

Rules for 0 - contains 2 rule(s)

Rule 1 for 0.0 (35, 0.973)

if Region <= 2 then 0.0

Rule 2 for 0.0 (6, 0.875)

if Employee <= 14 and Region <= 3 and TeleScale > 0 then 0.0

Rules for 1 - contains 3 rule(s)

Rule 1 for 1.0 (6, 0.875)

if Employee > 14 and Employee <= 37 and Region > 2 and Region <= 3 and TeleScale > 0 then 1.0

Rule 2 for 1.0 (12, 0.857)

if Employee <= 37 and Region > 3 then 1.0

Rule 3 for 1.0 (3, 0.8)

if Employee <= 16 and Region > 2 and TeleScale <= 0 then 1.0

Rules for 2 - contains 1 rule(s)

Rule 1 for 2.0 (65, 0.343)

if Region > 2 then 2.0

Rules for 3 - contains 1 rule(s)

Rule 1 for 3.0 (14, 0.5)

if Employee > 52 and Region > 2 and TeleScale <= 2 then 3.0

Rules for 4 - contains 1 rule(s)

Rule 1 for 4.0 (5, 0.714)

if Employee > 70 and Region > 2 and TeleScale <= 1 then 4.0

Default: 0

**Long-haul travel**

Rules for 0 - contains 2 rule(s)

Rule 1 for 0.0 (70, 0.986)

if Region <= 3 then 0.0

Rule 2 for 0.0 (19, 0.952)

if Employee <= 25 and TeleScale > 0 then 0.0

Rules for 1 - contains 2 rule(s)

Rule 1 for 1.0 (4, 0.833)

if Employee > 44 and Employee <= 80 and Region > 3 and TeleScale > 1 then 1.0

Rule 2 for 1.0 (30, 0.312)

if Region > 3 then 1.0

Rules for 2 - contains 2 rule(s)

Rule 1 for 2.0 (2, 0.75)

if Employee > 20 and Employee <= 44 and Region > 3 and TeleScale <= 0 then 2.0

Rule 2 for 2.0 (5, 0.571)

if Employee > 80 and Region > 3 then 2.0

Rules for 3 - contains 1 rule(s)

Rule 1 for 3.0 (17, 0.421)

if Employee > 44 and Region > 3 then 3.0

Default: 0