Frontiers of Engineering Management
DOI

ENGINEERING MANAGEMENT THEORIES AND METHODOLOGIES

Dong-ling Xu, Chris Foster, Ying Hu, Jian-bo Yang

Decision Support System for Evaluating Impact of Product Carbon Labeling Scheme

Abstract In this paper a decision support system for systematically evaluating the impact of labeling products with their carbon footprints is developed and applied to prioritize products for carbon labeling in a large supermarket chain in the UK. Carbon labels may change consumers' behavior and encourage suppliers to implement carbon-reduction solutions. Those changes may, however, lead to unintended risks. To handle the challenges of uncertainties in the evaluation, the Evidential Reasoning approach and the Intelligent Decision System software for multi-criteria decision analysis are applied to support the process. The system developed can be applied to assessing the impact of sustainable development policies to maximize their benefits and minimize their risks.

Keywords: impact assessment, multiple criteria decision analysis, risk analysis, carbon footprint, decision support system, product carbon labeling

1 Introduction

Research shows that there is increasing awareness among consumers to environment protection and climate change issues (Jacobsen, 2011). UK Carbon Trust was setup in 2001 and since 2006 has work with businesses through the use of product carbon labels to engage consumers and the various actors in product supply-chains in taking action to reduce the climate impacts of consumption. Various programs have been launched by manufacturers, retailers and labeling organizations, including PEPSICO, The Co-Operative, and Tesco etc. (Carbon Trust, 2008). A review by Bolwig and Gibbon (2009) identified fourteen product carbon footprinting schemes.

Carbon labels as under consideration in the UK are labels which convey to the reader all or part of the "carbon footprint" of the labeled product. The carbon footprint comprises the sum of greenhouse gas emissions arising throughout the production, distribution, use and disposal of a product (the product life cycle) expressed in terms of carbon dioxide equivalent (CO_2e). It is anticipated that carbon labels will initiate shifts of consumers towards buying lower carbon products and manufactures towards sourcing lower carbon raw material and technologies.

As well as opportunities to reduce the environmental impacts of consumption, there are risks associated with carbon labeling. For example, shifts of consumer purchases towards lower-value-added products have the unintended consequences of reduced profitability, and shifts towards sourcing low carbon but scarce raw materials cause unintended damage to ecosystems or societies. The risks can be grouped together as "unintended consequences". The mixed effects of a single issue campaign supported by an eco-labeling program are illustrated in Teisl, Roe, and Hicks (2002).

There are also direct costs associated with the carbon labeling activity itself. Aware of these costs, risks and opportunities retailers and labeling organizations seek a better understanding of them, firstly to estimate possible consequences before they actually happen and secondly to allow carbon labeling efforts to be directed towards groups of products for which the opportunity is greatest and the risks least. The aim of this paper is to provide systematic support to those retailers and organizations in such endeavors. Based on literature and observation, we first identify and explain the potential opportunities and risks that the product carbon labels may introduce. Those identified factors are then structured to form an assessment framework. Several groups of products are selected for comparison study so that the product groups that achieve high carbon reduction opportunities and low risks can be identified. We then further implement the framework into a decision support tool which can be used to by a large retailer in the UK to

(1) Prioritize product groups for the labeling program so that risks can be kept low and opportunity high

(2) Explore the risks and opportunity of labeling a product group in details

(3) Manage the knowledge and information obtained

Manuscript received December 5, 2013; accepted March 5, 2014

Dong-ling Xu (✉) Chris Foster, Ying Hu, Jian-bo Yang
Manchester Business School, The University of Manchester, Manchester, M15 6PB, UK
Email: L.Xu@mbs.ac.uk

from the process for supporting future decision making of similar nature

To deal with the uncertainty and diversity of information, we propose to apply a multi-criteria decision analysis (MCDA) approach, the Evidential Reasoning approach (Wang, Yang, & Xu, 2006; Yang & Xu, 2002), to structure the assessment framework, assess and compare the opportunities and risks of labeling different product groups. MCDA is a discipline in the decision sciences that consists of a family of concepts, procedures, methods and tools for identifying and structuring decision problems, conducting trade-off analysis among conflicting criteria and ranking different options based on a comprehensive assessment of them across the criteria (Belton & Stewart, 2002; Keeney & Raiffa, 1993). The ER approach is developed for handling both qualitative and quantitative information with hybrid uncertainties in complex decision problems. It uses belief distributions and belief decision matrices for problem modeling (Yang & Xu, 2002), which brings together in the same "assessment space" information of different kinds, qualitative and quantitative, deterministic and probabilistic, and complete and incomplete, relevant to a particular decision using a unified structure. The approach permits the incorporation and rational aggregation of as much data and other evidence as is available with judgments or expert opinions, as demonstrated by its applications in various other challenging assessment problems such as offshore structure safety assessment, environmental impact assessment, and risk assessment of nuclear waste repository options (Sii, Wang, Eleye-Datubo, Yang & Liu, 2005; Wang et al., 2006; Xu, Yang, Carlé, Har& deman, & Ruan, 2008).

The results presented in the paper are part of the outcomes generated from a study funded by the retailer. is used to support the analysis.

In addition to this introduction, this paper has the following main sections. In Section 2, the assessment criteria, or the components of both the potential opportunity and risks of a labeling program, are discussed. A MCDA model is introduced which is intended to provide a preliminary screening method for both data and judgment-based prioritization analysis. In Section 3, the process for defining and selecting product groups is discussed and examples of product categories selected are given. After a brief introduction to the ER approach in Section 4, the framework and process of conducting multiple criteria impact assessment of carbon labeling is illustrated in Section 5 using example product categories and data collected during the study. The paper is concluded in Section 6 by summarizing the main findings, recommendations, future work, and potential applications and limitations of the framework and process.

2 Impact assessment criteria hierarchy

An opportunity-risk analysis has, by definition, only two principal dimensions. The components of both dimensions are discussed below. Considering the challenges noted earlier, the first step in this study is to decompose the opportunity and risks described into their constituent elements that lend themselves easier to evaluate. Those elements are normally referred to as attributes or criteria, and they are used interchangeably in the paper.

2.1 The nature of the carbon reduction opportunity

The primary opportunity that the retailer and others wish to seize through carbon labeling is some reduction in greenhouse gas emissions. For a retailer, there may be other opportunities associated with the implementation of carbon-labeling – enhancement of reputation, growth in sales to green-minded consumers– but the "unique selling point" of carbon labeling is its potential to lead to lower greenhouse gas emissions. For organizations such as the Carbon Trust or the Department for Environment, Food and Rural Affairs (DEFRA) of the UK, whose missions are more focused on development of a lower-carbon economy than on improved financial results, this must be the driving force behind their development efforts. Therefore, it is natural that an evaluation of the anticipated benefits of a carbon labeling scheme should focus on the magnitude of this potential emissions reduction. To describe the scale of this hoped-for benefit, the term "Carbon Opportunity" is coined in the study, while the risk dimension is termed as "non-Carbon risk".

2.2 Components of carbon opportunity

Reduction in greenhouse gas (GHG) emissions will only follow a carbon-labeling initiative if some change occurs. These changes could take the form of either a shift on the part of (some) consumers to purchase "lower-carbon" products to fulfill a particular need or emissions reduction somewhere in the value-chain of the labeled product. Therefore "carbon opportunity" has two principal components:

(1) Product-shift related opportunity comprises GHG emission reductions linked to the switch by consumers from buying products with higher carbon footprints to buying those with lower ones. A change in the relative sales volume of different products is implied.

(2) Improvement related opportunity comprises potential reductions in the carbon footprint of individual products. This has been considered from the perspective of changes made by upstream actors in the supply chain of particular products. These changes might affect GHG emissions at any stage of the life cycle; changes in raw material specification affect emissions at the primary production stage, while changes in formulation can affect emissions in use (e.g. laundry detergents formulated to enable lower washing temperature could offer major improvements in overall carbon footprint).

This distinction also facilitates further analysis of the potential scale of the carbon opportunity. The following considerations quickly arise:

(3) If the product shift-related opportunity is to be esti-

mated, it is necessary to make judgments about which products consumers might actually switch between when faced with a carbon label. We term these products "substitutable products".

(4) Information about the carbon footprints of these different products is also needed. Here the differences instead of the absolute values between the carbon footprints of substitutable products are important.

(5) Supplier-improvement is decided by whether suppliers are willing and able to change. Thus information about suppliers is required for reliable evaluation.

(6) The amounts of substitutable products sold (in terms of number, weight or volume) also clearly have a strong bearing on the scale of the opportunity.

2.3 Elements of non-carbon risk

Having considered what the components of carbon opportunity are, we next define the components of non-carbon risk. Keeping to the very broad level, any discussion of the implications of change for the long-term sustainability of human society in its current form will identify the need to consider impacts on the environment, people and social structures ("society"), and the economy. So non-carbon risk is likely to have economic, environmental and social components.

Further consideration of risks takes into account the fact that it is a retailer that is making the decision in this case, and therefore the perspective of a retailer on the relevance of various risk components must be incorporated. Discussion with the retailer's personnel has informed the selection of risk components set out below.

2.3.1 Economic risk

In this realm, any retailer undertaking a program such as carbon labeling is inevitably concerned with the potential impacts of the program on its own profitability. Will margins across the product group increase or decrease, and by how much? By how much will sales measured in financial terms fall or rise? How far will the retailer's sales in general fall or rise as a result of the implementation of carbon labeling?

For clarity, we termed this commercial risk since it largely ignores possible impacts on the wider economy.

The third question reflects the possibility that introducing carbon-labeling of products might cause some shoppers to shift from the retailer of the study to other retailers. The perceived driving force being the potential for the messages given by carbon labels about groups of products to confuse consumers (who are already bombarded with what could generally be termed "product-related messages" when in a large shop), or run counter to their existing belief, to the extent that they cease to have the "trust" in the company that it considers vital to successful and profitable customer retention.

2.3.2 Environmental risk

The risks to the environment associated with a change in patterns of production and consumption can be divided into two:

(1) Pollution risk is the risk associated with non-greenhouse gas pollutant increases as change takes place in supply-chains or on the part of consumers. An example is particulate emissions increases which could result from motorists switching from petrol- to diesel-fuelled cars, with consequent effects on human and animal health. Since air pollution, water pollution and land pollution can have different sources, separate consideration of the following three elements is necessary:

① Air pollution risk, associated with changes in non-greenhouse gas emissions to air throughout the product life cycle.

② Water pollution risk, associated with changes in emissions to water throughout the product life cycle.

③ Land pollution/waste risk, associated with changes in emissions to land – including deposits of wastes - throughout the product life cycle. Wastage also represents inefficient utilization of inputs, even when it undergoes some forms of recovery.

(2) Resource risk is the risk that resources, including water resources, mineral deposits and the resources embodied in certain types of land, are depleted more quickly as a result of changes made to seize the carbon opportunity, or that these changes cause producers to switch from more sustainable to less sustainable reserves as sources of primary inputs. A switch to biofuels reducing greenhouse gas emissions at the expense of the depletion of biological resources embodied in unexploited rainforest is one example of resource risk.

Impact assessment within environmental life cycle assessments (LCAs) tends to treat land use, water use, biotic resource depletion and abiotic resource depletion as four categories of resource-related impact (see, e.g. Udo de Haes,, Finnveden, Goedkoop, Hauschild, Hertwich, & Jolliet, 2002). Accepting that the abiotic and biotic resource depletion is driven by material and fuel selection, while water use has somewhat different drivers, we simplify resource risk into two components at this early stage of analysis:

(1) Raw material risk, being the risk that changes in the pattern of raw material use following a carbon labeling exercise lead to the more rapid depletion of natural resource reserves

(2) Water risk, being the risk that changes following a carbon labeling exercise increase pressure on water resources.

2.3.3 Risks to wider society (ethical risk)

Beyond the environment and retailers' profits, many examples of potential "unintended" consequences of changes following a carbon labeling program can be identified. Some of these will be seen as undesirable by one or more groups in society. If fur jackets turn out to be lower-carbon than coats made from wool or synthetic fibers, animal welfare may suffer; if hothouse techniques produce low-carbon cocoa, farmers in developing countries may lose their markets; and if high-fat, sugary confectionery is the lowest-carbon source of calories available, some health-promotion messages may be at odds with GHG emission reduction. Because the protection of consumers' health, animal welfare and the rights of workers in supply-chains has come to be perceived as a moral duty for large-scale purchasers such as major retailers, we grouped these risks together under the heading "Ethical risk".

Clearly ethical risks are closely linked to the structure of supply-chains, and the locations and behaviors of actors in those supply-chains. Further analysis suggested that ethical risk could be considered as linked to changes in one or more of four categories of impact:

(1) Impacts on society local to production activities.

(2) Impacts on employees in businesses in the supply-chain.

(3) Impacts on consumers' health.

(4) Impacts on animals.

The last of these is clearly not relevant to groups of products in which animal-derived materials play no part.

To summarize, the criteria for assessing the impact of a carbon labeling program should have the components as structured in the hierarchy shown in Figure 1.

3 Identifying product groups for carbon labeleing decision and data collection

In section 2.2, it was noted that one of the two principal components of carbon opportunity is the product shift related opportunity. It exists only if a number of substitutable products carry carbon labels. Therefore a good candidate for carbon labeling is a group of products which are seen by a reasonable proportion of consumers as being substitutable.

This has implicitly been recognized in the groups of products for which the retailer in this study has decided to test carbon labeling: kettles, laundry detergents and light bulbs, for example. These groups do not all correspond to single sub-groups in its publicized hierarchy of products and categories, but clearly some judgment (whether based on "common sense" or sales data) led to them being considered as "product groups" for this purpose.

Deciding what is substitutable is of course a matter of subjective judgment. For instance, one individual seeking bread may be willing to buy any brown or white loaf of the required size, and another may be only willing to buy one type of brown loaf. To permit this analysis to be conducted in a practical timeframe, the number of groups must however be manageable. This requires a rather liberal view about which products might be substitutable to be taken. If the extreme opposite view were adopted, i.e. that no two products represent substitutes for each other, then the notion that carbon labels might lead consumers to change their behavior becomes questionable.

The nature of the retailer's publicized hierarchy of product categories, groups and sub-groups makes it a rather unsatisfactory starting point for grouping products for carbon-labeling. Some sub-groups relate only to a single brand, whilst others are highly heterogeneous or relate to particular types of promotion. So rather than screening thousands or so product sub-groups on the basis of their substitutability, they have been used as the basis for creating a smaller number of

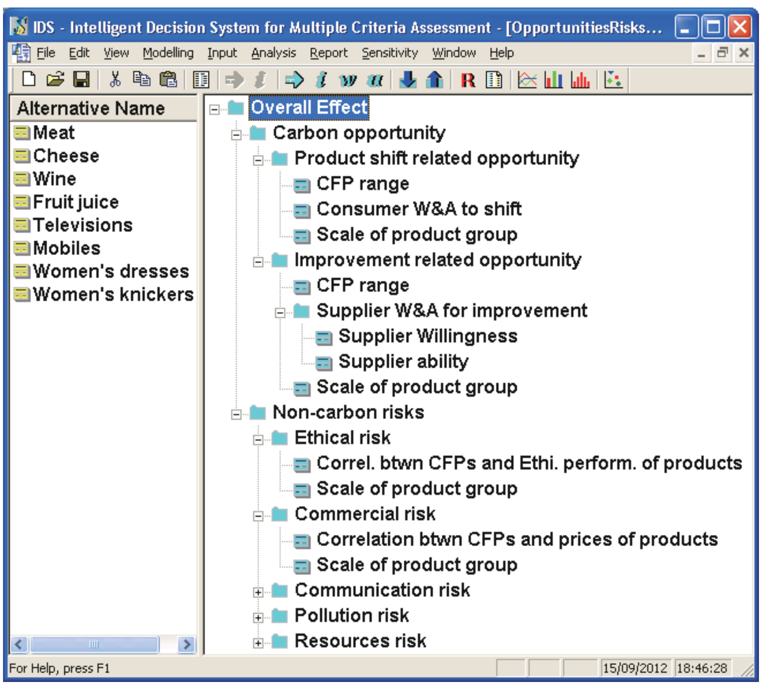

Figure 1. Impact assessment criteria hierarchy and product groups to be assessed.

groups which do lend themselves to this assessment, and to further analysis in this study.

Starting from the product sub-groups designated in the retailer's own system, groups were defined on the basis of a combination of the function provided by the products in the groups and the characteristics of those products – these characteristics reflected either aspects of the products' make up or the way in which they are used. This exercise first places each of the retailer's product sub-categories into one of 18 general categories; these general categories are subdivided (e.g. the category "baby" is broken down into "baby food and drink" and "baby other") either once or twice (e.g. "baking" is a group within the general category "food" but "baking" is further divided into "sugar", "flour/flour-based", and so on). Given the timescale of this project, the priority in this classification exercise was identification of a number of manageable product groups for further analysis. Some considerations for selecting products and product categories includes

(1) Product availability across brands from multiple suppliers.

(2) Likelihood of customer's shift from one product to another within a category.

(3) Intuitive low risk of labeling products in a category.

(4) Supplier willingness to work with a retailer in a category.

(5) Easiness for customers to change and compare product differences.

(6) Product sales volume.

Eight product groups are identified for validation study. They are meat, cheese, wine, fruit juice, TVs, mobile phones, women's dresses and women's knickers (see Figure 1). In each of these eight categories, five to seven products are selected for data collection and comparative study.

Data collected for those eight selected product groups include information provided by the retailer such as sales volume, carbon dioxide emission estimates based on information identified in literature, manufacturing process of the products or consulting experts in the product production. Examples of the data collected are shown in the examples in Section 5.2. Four MSc students worked for the project for three months under the supervision of staff members with expertise in product life cycle assessment CO_2 emission, and decision analysis and support.

4 Data aggregation method

The assessment and comparison of opportunities and risks associated with labeling different products groups involve the consideration of a diverse range of often conflicting assessment criteria and different alternatives (or groups of products) to be assessed and ranked. This type of problems is known as multiple criteria decision analysis (MCDA) problems (Keeney & Raiffa, 1993). Over the past 40 years, with the advancement of computer technology, many methods and tools have developed for modeling and analyzing MCDA problems, such as those reviewed by Hwang and Yoon (1981), French and Xu (2005) and Figueira, Greco, and Ehrgott (2005). Many of the methods and tools are based on modeling MCDA problems using average numbers and decision matrix (Hwang & Yoon, 1981) which has been criticized as lack of transparency and are unable to explicitly represent and analyze uncertainty (Savage, 2009) in MCDA problems. The problem we have in this study, however, has not just one but hybrid types of uncertainty:

(1) Uncertainty about the nature of the anticipated changes themselves

"Changes in consumer purchasing patterns" and "Improvement in the supply-chain" are mentioned as potential responses to carbon labeling. Although examples of the use of carbon footprints for supply-chain improvement can be found (see e.g., the Walker's Crisps case study found at http://www.walkerscarbonfootprint.co.uk/walkers_carbon_trust.html and last accessed 15th Sept 2010), at this early stage the extent of potential benefits is uncertain while some research has found the role of environmental information in tackling environmental impacts to be limited, in the food supply chain at least (Fuentes & Carlsson-Kanyama, 2006).

(2) The diversity of potential risk factors and the diversity of available methods for evaluating each of them

(3) Unknown or partially known information of both carbon footprints and other social or environmental impacts arising from the production and use of those products

(4) Subjective judgments

Faced with challenges like this and decisions of a similar nature, many decision analysts and decision makers resort to highly qualitative, heavily judgment-based approaches to evaluation.

To analyze this challenging problem, we propose to apply an advanced MCDA approach, the Evidential Reasoning approach (Yang & Xu, 2002), an evolving approach for analyzing MCDA problems under hybrid uncertainties.

4.1 The evidential reasoning (ER) approach for multiple criteria decision analysis (MCDA)

There were calls in early 1990s to develop new methods that could produce consistent and rational results, be capable of dealing with uncertainty and providing transparency to the analysis processes (Dyker, Fishburn, Steuer, Wallenius, & Zionts, 1992; Stewart, 1992). In answering those calls, the ER approach (Yang & Xu, 2002) for MCDA was initially proposed in the 1990s (Yang & Singh, 1994) and is still evolving (Yang & Xu, 2011). During the period, through both theoretical and applied research, significant efforts and progress have been made in investigating and validating the rationality and reliability of the ER approach in handling both qualitative and quantitative information, and data with hybrid uncertainties such as inaccuracy, incompleteness and randomness in MCDA problems.

Using the concept of belief distributions and belief decision matrix, as illustrated in the following three sub sections, the ER approach models uncertainty in complex decision problems explicitly and analyzes its effect on decision outcomes in a structured, systematic and consistent manner. The belief distributions overcome the flaws of averages and improve transparency and informative-ness. The approach is implemented in a software package, Intelligent Decision System (IDS) (Xu, McCarthy & Yang, 2006), and examples of its previous applications include: environmental management (Wang et al., 2006; Xu et al., 2008), risk assessment (Kong, Xu, Body, Yang, Mackway-Jones, & Carley, 2012; Wang et al., 2004), safety assessment (Liu, Yang, Ruan, Martinez, & Wang, 2008), consumer preference identification and new product development (Chin, Yang, Lam, & Guo, 2009; Maddulapalli, Yang, & Xu, 2012; Yang, Xu, Xie, & Maddulapalli, 2011).

Before the application of the ER approach and the implementation of an assessment tool for analyzing the impact of the carbon label scheme, the problem modeling and information aggregation processes of the ER approach are illustrated using simple examples (with probability and missing data) in the following three sub sections.

4.2 Belief decision matrix and assessment problem modeling in the ER approach

Suppose an assessment problem has M alternatives (or product groups in this study) to be assessed on L criteria. A decision matrix is widely employed to record the assessment information, as shown in Table 1, where a_{ml} is a single number, normally an average score given to Alternative m assessed on Criterion l.

Table 1 Decision Matrix and Belief Decision Matrix

	Criterion 1	…	Criterion l	…	Criterion L
Alternative 1	a_{11}		a_{1l}		a_{1L}
…					
Alternative m	a_{m1}		$\mathrm{a_{ml}}$		a_{mL}
…					
Alternative M	a_{M1}		$\mathrm{a_{Ml}}$		$\mathrm{a_{ML}}$

In the ER approach, the problem is modeled through a belief decision matrix, where each element in the matrix is a belief or probability distribution (Yang & Xu, 2011), instead of an average score. Let

$$H=\{H_1, \dots, H_N\} \tag{1}$$

be a collectively exhaustive and mutually exclusive set of assessment grades (or scores) where N is the number of grades in the set. Then each element, such as a_{ml}, in a belief decision matrix can be expressed as

$$a_{ml}=\{(H_1, \beta_{l,1}),...,(H_N, \beta_{l,N}), (H_{1N}, \beta_{l,1N})\} \tag{2}$$

Where $0\leq\beta_{1,i}\leq1$(i=1, ... N; l=1,..., L) is a belief degree to which the performance of the alternative is assessed to the grade H_i on criterion l, $\beta_{l,1N}=1-\sum_{i=1}^{N}\beta_{l,i}\geq0$ and H_{1N} is the grade interval from H_1 to H_N. H_{1N} is used to accommodate unknown information or uncertain judgements. An assessment expressed in the format of equation (2) is referred to as a belief distribution or a performance (belief) distribution. When $\sum_{i=1}^{N}\beta_{l,i}=1$ or $\beta_{l,1N}=0$, the assessment is said to be complete (no missing information), otherwise ($\sum_{i=1}^{N}\beta_{l,i}<1$ or $\beta_{l,1N}>0$) incomplete.

For example, suppose that we use three grades, Low, Average, and High, to measure Product Shift Related Opportunity (the 1st sub criterion of Carbon Opportunity) of labeling TVs group, that is $\{H_1, H_2, H_3\}$={Low, Average, High}. Further suppose that the outcome of the assessment is the following distribution

$$a_{TVs,1}=\{(H_1, 30\%), (H_2, 50\%), (H_3, 20\%)\} \tag{3}$$

Similarly for other criteria, such as Improvement Related Opportunity (the 2nd sub criterion of Carbon Opportunity) of labeling TVs group, suppose the assessment outcome is represented as the following distribution

$$a_{TVs,2}=\{(H_1, 40\%), (H_2, 30\%), (H_3, 20\%)\} \tag{4}$$

The belief degrees in the distributions can be viewed as probability or frequency, which can be estimated from evidence or judgments. They can also be obtained from combining the evidences collected and judgments made for the alternative on sub criteria. A belief distribution can model different types of uncertainty in an evaluation of a performance and enable their effects on the assessment outcomes to be explicitly shown without resorting to sensitivity analysis (Xu, Yang, & Wang, 2006). The types of uncertainty the belief structure can represent include: randomness, subjective judgments, and partial or complete missing data.

To facilitate comparison, the evidences and judgments made for an alternative against sub criteria need to be aggregated. The ER aggregation algorithm (Yang & Xu, 2002) can be applied for this purpose. The aggregation process including the handling of missing data is given in the following sub sections..

4.3 Information aggregation in the ER approach

The information aggregation algorithm (Yang & Singh, 1994; Yang & Xu, 2002) in the ER approach is developed based on evidence theory (Shafer, 1976). Evidence theory can be regarded as an extension to Bayes theory, plays an import role in uncertainty analysis (see French, 1988, p.254) and has been widely used in many areas including decision analysis and artificial intelligence to handle uncertain information (Beynon, 2005).

The inputs to the algorithm are the performance distributions of an alternative on all the lowest level criteria in the criteria hierarchy, or all the elements in a row in Table 1. The outcome of the algorithm is also a distribution, a_m, which is normally referred to as the overall performance distribution of the alternative. That is

$$a_m=\{(H_1, \beta_1),\dots (H_N, \beta_N), (H_{1N}, \beta_{1N})\} \\ =\omega_1 a_{m1}\oplus\dots\oplus\omega_l a_{ml}\oplus\dots\oplus\omega_L a_{mL} \tag{5}$$

where $\omega_1,\dots,\omega_L$ are weights of criterion 1,…,L respectively and ⊕ means orthogonal sum. Details of the calculations are illustrated using a numerical example in the next sub section. The algorithm is able to preserve the richness of information in performance distributions during the aggregation process and provide statistically meaningful lower and upper bounds of the effects of any partial or complete missing data (Yang & Xu, 2002). More detailed theoretical and general discussions of the algorithm can be found in Yang and Xu (2002). The application of the algorithm is facilitated by the IDS software tool (Xu et al, 2006) and some of its interfaces are shown as figures in the paper.

Sometimes it is desirable to aggregate the performance distribution of equation (5) further into a score for directly ranking alternatives. Suppose the utility (Yang, 2001; see also Goodwin & Wright, 1999, p.104) of the grade H_i is $u(H_i)$. Usually $u(H_i)$ is a value between 0 and 1, with 1 indicating the most preferred and 0 the least preferred outcome respectively. Without losing generality, suppose H_1 is the most preferred and H_N the least. Using the utilities, the aggregated performance score of the m-th alternative can be calculated from the distribution of equation (5) as

$$u(a_m)=\sum_{i=1}^{N}\beta_i u(H_i) \tag{6}$$

If there is unknown information in the performance distribution, the upper and lower bounds of the score, $u_{\max}(a_m)$ and $u_{\min}(a_m)$ respectively, can be calculated as (Yang, 2001)

$$u_{\max}(a_m)=\sum_{i=1}^{N-1}\beta_i u(H_i)+(\beta_N+\beta_{1N})u(H_N) \tag{7}$$

$$u_{\min}(a_m)=(\beta_1+\beta_{1N})u(H_1)+\sum_{i=2}^{N}\beta_i u(H_i) \tag{8}$$

The average of the two is $u_{avg}(a_m)$ with

$$u_{avg}(a_m)=\frac{u_{\max}(a_m)+u_{\min}(a_m)}{2} \tag{9}$$

The average score, $u_{avg}(a_m)$, can be used as an indicative score for ranking alternatives, but it should be noted that it is just the middle point in the range in which the utility score of the alternative will fall, depending on what the missing data turn out to be.

4.4 Numerical example of the ER aggregation algorithm

In this sub section, the orthogonal sum in equation (5) is explained using a simple example. The calculation may initially look complicated, but the process is well structured and easy to be implemented into software tools. It is implemented in both IDS and Excel Spreadsheet.

Suppose that we wish to know the potential "Carbon opportunity" of labeling the product group of TVs, we need to aggregate its performance on "Product shift related opportunity" shown by equation (3) and "Improvement related opportunity" shown by equation (4). In another word, we need perform $\omega_1 a_{TVs,1}\oplus\omega_2 a_{TVs,2}$ where ω_1 and ω_2 are weights of the corresponding criteria reflecting their relative importance. If there are more than two criteria to be aggregated, the third one can then be aggregated with the orthogonal sum of the first two, and so on. The order of the aggregation does not affect the final outcome (Shafer, 1976).

Suppose the importance weights for the two sub criteria are 0.3 and 0.2 respectively. In the ER algorithm, the weights of sub-criteria need to be normalized to 1. Suppose there are L sub-criteria and ω_i is the weight of the i-th sub-criterion (i=1,…,L), then the normalized weights should satisfy

$$0\le\omega_i\le 1 \tag{10}$$

and

$$\sum_{i=1}^{L}\omega_i=1. \tag{11}$$

In the example, the normalized weights for sub-criteria 1 and 2 should be ω_1=0.3/(0.3+0.2)=0.6 and ω_1=0.2/(0.3+0.2)=0.4 respectively.

The ER aggregation algorithm takes the two assessments $a_{TVs,1}$ and $a_{TVs,2}$ as its inputs and generate a combined assessment, denoted by $a_{TVs,Oppt}=\omega_1 a_{TVs,1}\oplus\omega_2 a_{TVs,2}=\{(H_1,\beta_1),(H_2,\beta_2),(H_3,\beta_3)\}$, as its output. The belief degrees in the combined assessment, β_1, β_2 and β_3, are obtained through the following steps:

Step 1: Calculate basic probability masses p_n and q_n associated with grade H_n (n=1, 2, 3), and $\bar{p}_H$, $\tilde{p}_H$, $\bar{q}_H$, and $\tilde{q}_H$:

$$p_n=\omega_1\beta_{n,1}\ (\mathrm{n}=1,2,3) \tag{12a}$$

$$\bar{p}_H=1-\omega_1=0.4 \tag{12b}$$

$$\tilde{p}_H=\omega_1(1-\sum_{n=1}^{3}\beta_{n,1})=0 \tag{12c}$$

with

$$p_H=\bar{p}_H+\tilde{p}_H=1-\omega_1\sum_{n=1}^{3}\beta_{n,1}=1-\omega_1=0.4 \tag{12d}$$

$$q_n=\omega_2\beta_{n,2}\ (n=1,2,3) \tag{13a}$$

$$\bar{q}_H=1-\omega_2=0.6 \tag{13b}$$

$$\tilde{q}_H=\omega_2(1-\sum_{n=1}^{3}\beta_{n,2})=0.4\times0.1=0.04 \tag{13c}$$

with

$$q_H=\bar{q}_H+\tilde{q}_H=1-\omega_2\sum_{n=1}^{3}\beta_{n,2}=1-\omega_2\times0.9$$
$$=1-0.36=0.64 \tag{13d}$$

where $\beta_{n,1}$ and $\beta_{n,2}$ are the belief degrees associated with H_n (n = 1, 2, 3) in equation (3) and equation (4) respectively. For example, $\beta_{1,2}$=0.3 and $p_1=\omega_1\beta_{1,1}$=0.6×0.3=0.18. The calculated p_n and q_n are called probability masses assigned to grade H_n. The terms p_H and q_H in equations (12d) and (13d) are the remaining probability masses initially unassigned to any individual grades. The term p_H consists of two parts, $\bar{p}_H$ and $\tilde{p}_H$, as shown in equation (12d). The first part $\bar{p}_H$ represents the degree to which other criteria can play a role in the assessment. It should eventually be assigned to individual grades in a way that is dependent upon how all criteria are weighted and assessed. The second part, $\tilde{p}_H$=0, indicates that the amount of unknown information in an assessment which is 0 (percent) here because $\sum_{n=1}^{3}\beta_{n,1}$=1(see equation (12c)).

Similarly, q_H consists of two parts, $\bar{q}_H$ and $\tilde{q}_H$. Note that $\tilde{q}_H$ is not zero due to the incompleteness of the assessment $a_{TVs,2}$, or $\sum_{n=1}^{3}\beta_{n,2}$=0.9. The parts $\tilde{p}_H$ and $\tilde{q}_H$ represent the remaining probability mass unassigned due to the incompleteness in their corresponding original assessments. They are proportional to their corresponding missing belief degrees and criterion weights, and will cause the subsequently aggregated assessments to be incomplete (Yang & Xu, 2002). The values of those calculated probability masses are given in the 1st row and 1st column of Table 2.

Step 2: Calculate combined probability masses and combined belief degrees β_1, β_2 and β_3:

The above probability masses are aggregated into the following combined probability masses, denoted by r_n (n = 1, 2, 3), $\bar{r}_H$ and $\tilde{r}_H$, using the following equations:

$$r_n=k(p_n q_n+p_H q_n+p_n q_H),\ (n = 1, 2, 3) \tag{14}$$

$$\bar{r}_H=k(\bar{p}_H \bar{q}_H) \tag{15}$$

$$\tilde{r}_H=k(\tilde{p}_H \tilde{q}_H+\tilde{p}_H \bar{q}_H+\bar{p}_H \tilde{q}_H) \tag{16}$$

where

$$k=\left(1-\sum_{t=1}^{3}\sum_{\substack{n=1\\ n\neq t}}^{3} p_t q_n\right)^{-1} \tag{17}$$

From Table 2, we have

k=(1-(0.0192+0.048+0.0144+0.0216+0.024+0.0144))$^{-1}$ =0.8584^{-1}=1.1650

$r_1=k(p_1q_1+p_Hq_1+p_1q_H)$=1.1650×(0.0288+0.064+0.108 +0.0072)=0.2423

$r_2=k\times(p_2q_2+p_Hq_2+p_2q_H)$=1.1650×(0.036+0.048+0.18+0.012) =0.3215

$r_3=k\times(p_3q_3+p_Hq_3+p_3q_H)$=1.1650×(0.0096+0.032+0.072 +0.0048)=0.1379

$\tilde{r}_H=k(\bar{p}_H\bar{q}_H)$=1.1650×0.24=0.2796

$r_H=k(\tilde{p}_H\tilde{q}_H+\tilde{p}_H\bar{q}_H+\bar{p}_H\tilde{q}_H)$=1.1650×0.016=0.0186

If there are more than two sub-criteria, the combined probability masses can then be combined with the probability masses of the performance distribution on the 3rd criterion in the same way. The process is repeated until all the sub-criteria are combined. If there are several levels in a criteria hierarchy, the aggregation process is carried out from the bottom level until the top of the hierarchy is reached.

The belief degrees β_n (n = 1, 2, 3) in the aggregated performance distribution

$$a_{TVs,Oppt.}=\omega_1 a_{TVs,1}\oplus\ldots\oplus\omega_2 a_{TVs,2}\quad \{(H_1,\beta_1),(H_2,\beta_2),(H_3,\beta_3)\} \tag{18}$$

are calculated from the combined probability masses by:

$$\beta_n=\frac{r_n}{1-\bar{r}_H} \tag{19}$$

For the example, they are given by

$\beta_1=\dfrac{0.2423}{1-0.2796}=0.3364,\ \beta_2=\dfrac{0.3215}{1-0.2796}$ $=0.4463$, and β_3=0.1915.

Because of the incompleteness in one of the assessments, the aggregated assessment is also incomplete and it is shown by the sum of the three belief degrees which is 0.9741, indicating only 97.41 % of the belief degrees are assigned. The unassigned part is given by,

β_H=1−0.9741=0.0259

It could be partially or completely assigned to any combination of the three grades depending on what information is in the missing data. If it were assigned to grade H1, the belief degree associated to this grade could be as high as $P(H_1)=\beta_1+\beta_H$=0.3623. Similarly, it could also be assigned to grades H2 and H3. Therefore β_H represents the combined effects of missing data on aggregated outcomes.

If necessary, a utility score can be calculated from the aggregated assessment. Suppose the utilities for the 3 grades are (H_1, H_2, H_3)=(0, 0.5, 1). The expected utility score of the assessment given by equation (18), denoted by U, can be calculated as follows with the belief degrees as weights,

$$U=\sum_{i=1}^{3}u(H_i)\beta_i=0.4147 \tag{20}$$

Table 2 Probability Masses

	$P(S_1)$ ⊕ $P(S_2)$	$P(S_1)$: p_1=0.18 {H_1}	p_2=0.3 {H_2}	p_3=0.12 {H_3}	$\bar{p}_H$=0.4 {H}	$\tilde{p}_H$=0 {H}
$P(S_2)$	q_1=0.16 {H_1}	p_1q_1=0.0288 {H_1}	p_2q_1=0.048 {Φ}	p_3q_1=0.0192 {Φ}	$\bar{p}_Hq_1$=0.064 {H_1}	$\tilde{p}_Hq_1$=0 {H_1}
	q_2=0.12 {H_2}	p_1q_2=0.0216 {Φ}	p_2q_2=0.036 {H_2}	p_3q_2=0.0144 {Φ}	$\bar{p}_Hq_2$=0.048 {H_2}	$\tilde{p}_Hq_2$=0 {H_2}
	q_3=0.08 {H_3}	p_1q_3=0.0144 {Φ}	p_2q_3=0.024 {Φ}	p_3q_3=0.0096 {H_3}	$\bar{p}_Hq_3$=0.032 {H_3}	$\tilde{p}_Hq_3$=0 {H_3}
	$\bar{q}_H$=0.6 {H}	$p_1\bar{q}_H$=0.108 {H_1}	$p_2\bar{q}_H$=0.18 {H_2}	$p_3\bar{q}_H$=0.072 {H_3}	$\bar{p}_H\bar{q}_H$=0.24 {H}	$\tilde{p}_H\bar{q}_H$=0 {H}
	$\tilde{q}_H$=0.4 {H}	$p_1\tilde{q}_H$=0.0072 {H_1}	$p_2\tilde{q}_H$=0.012 {H_2}	$p_3\tilde{q}_H$=0.0048 {H_3}	$\bar{p}_H\tilde{q}_H$=0.016 {H}	$\tilde{p}_H\tilde{q}_H$=0 {H}

The score will normally be different from that calculated by using the weighted sum approach as the ER aggregation is a nonlinear process in which harmonic judgments will be reinforced more than proportionally and conflicting ones weakened accordingly. More details on the properties of the ER algorithm can be found in Yang and Xu (2002) and Yang (2001). In the example, we have assumed that the same set of grades is used for both sub criteria. When aggregating assessments made using different grade sets for different criteria, Yang (2001) has provided information transformation techniques to first convert the assessments into ones based on a common set of grades while preserve utility equivalence during the conversion.

5 Carbon label impact modeling and analysis using the ER approach

In the next two sub sections, the first one is focused on problem modeling - applying the ER approach and the IDS software to implement a tool that manages assessment knowledge and information and support the assessment of carbon label impact. The second one is on analysis and assessment outcomes - the applications and illustration of the tool, including functions and interfaces for exploring the effects of uncertainties on outcomes, and interactive graphical interfaces and text reports for decisions and risk communication.

5.1 Problem structuring and modeling using IDS

Problem structuring and modeling is a process to elicit, record and structure assessment knowledge and decision makers' preferences. This process is relatively independent of any particular product groups to be considered. Using IDS, the process has the following steps.

5.1.1 Build the assessment criteria hierarchy

The assessment criteria as discussed in Section 2 can be structured as the hierarchy shown in the right pane of the IDS main window (see *Figure 2*). The eight product groups are listed in the left pane.

As it is very difficult to estimate the exact carbon footprint for each product, we examine the patterns of shifts and changes of buyers' and suppliers' behavior within a group of products as a whole from a macroscopic point of view and propose to check the relative CFP (Carbon Footprint) levels among product in the same group and the relative risk levels of labeling them. This is done by further decomposing each of the opportunity and risk criteria into sub criteria.

Opportunity is assessed through three sub criteria: CFP Range (or difference between high and low carbon products) within a group, consumer or supplier W & A (Willingness and Ability) to Shift or change, and scale of product group in terms of sales volume. This consideration is based on the assumption that the larger the difference between the CFP of high and low carbon products within a group, the more willing and able customers (or suppliers) are to shift (change), and the higher the sales volume, the greater the opportunity.

Similarly each risk criterion is assessed through two sub criteria: correlation between the relative CFP level of a product and its corresponding risk level, and the scale of the product group in terms of sales volume. This is based on the assumption that if low (or high) carbon products are normally associated with low (or high respectively) level of risks within a product group and the sales volume of the group is

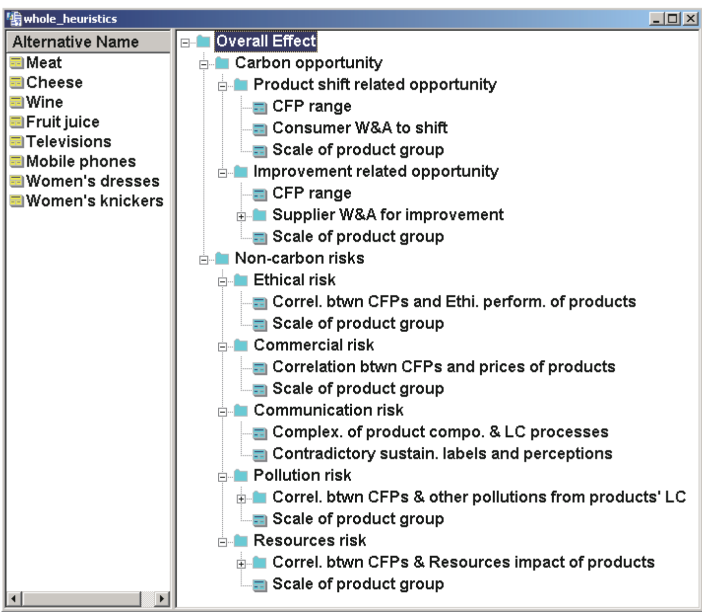

Figure 2. Extended criteria hierarchy for assessing impact of carbon labeling a product group. All figures in the paper are generated by IDS.

high, then the risk of labeling this group is low.

The criteria hierarchy including the sub criteria is shown in Figure 2.

5.1.2 Define each criterion

A criterion can be quantitative or qualitative in nature. For each criterion, whether qualitative or quantitative, a range that contains the best and worst possible outcomes needs to be identified, and a utility function (Keeney & Raiffa, 1993) which represents decision makers' preferences towards the outcomes in the range needs to be defined. A utility function maps the outcomes to a value between 0 and 1, with 1 associated with the most preferred and 0 the least preferred outcomes. The shape of the functions can be linear or nonlinear. Normally a linear utility function indicates a risk neutral attitude of a decision maker, while a concave or convex function, risk aversion or risk taking respectively.

To establish the shape of a utility function, normally a few referential points are identified in the range, and then each point or grade is assigned a utility. For example, for the top level criterion (Overall effect) and the second-level criteria (Carbon Opportunity and Non-carbon Risk), five grades and their utilities are suggested as shown in Table 3. The continuous utility function can be obtained by joining each two adjacent points with a straight-line to form a piecewise linear function. The overall shape of the function is linear in this case. In this study, utility functions for all other evaluation criteria are assumed to be linear to reflect a risk-neutral attitude of the retailer's management team.

Table 3 Utility Functions for the Assessment Grades of the Top and Second Level Attributes

Overall Effect Assessment grades	Carbon opportunity Assessment grades	Non-carbon risk Assessment grades	Utility
Best	Very high	Very low	1
Good	High	Low	0.75
Average	Average	Average	0.5
Bad	Low	High	0.25
Worst	Very low	Very high	0

5.1.3 Assign a weight to each criterion

Weights are assigned to assessment criteria to represent their relative importance. Weight assignment is normally subjective and there are different supporting methods built into the IDS tool to help reduce the subjectivity. If there are uncertainties in the weights assigned, there are various sensitivity analysis facilities in IDS to check the effects of the uncertainties and some are illustrate in Sections 5.2.6 and 5.2.8.

In this study, for all sub-criteria at the same level and sharing the same upper level criterion in the hierarchy, it is assumed that they have equal weights. Detailed study on whether non-equal weights are desirable is not conducted in this study due to resource constraints and can be a topic for further research.

5.2 Assessment, analysis and outcomes

Once the assessment model is implemented in IDS, it can then be used as a tool to support the assessment process. Data needs to be collected or estimated regarding each of the product groups on each of the lowest level criteria in the criteria hierarchy (see Figure 2). The data are then recorded using the tool, and aggregated using the ER approach built into it so that the overall impact against upper level criteria can be generated. As some of the data are difficult to find or estimate, uncertainty or missing information are present in the data. As illustrated in Section 4, the ER approach can model problem and aggregate information with various types of uncertainties. To avoid repetitiousness, we use the assessment of TVs and mobile phones against the criterion "Product shift related opportunity" (see *Figure 2*) to illustrate the process. In the following discussions, A, B, C, D, E, and F are six models of TVs or mobile phones, in their corresponding contexts, sold by the retailer.

5.2.1 Assessment product shift related opportunity

To illustrate application of the developed assessment tool and the process, the assessment of Product Shift Related Opportunity, one of the main criteria, is given in details in this section. This opportunity is related to the following three factors:

(1) Carbon footprint (CFP) range, based on the assumption that the larger the range, the more green house gas (GHG) emissions will be reduced as consumers shift from higher CFP products to lower CFP ones.

(2) Consumer willingness and ability to shift, based on the assumption that the more willing and able the consumers are to shift from higher CFP products to lower CFP one, the more GHG emissions will be reduced.

(3) Scale of product group in terms of sales volume, based on the assumption that the larger the scale, the more the shifts, and the more GHG emissions will be reduced.

The assessment of product-shift related opportunity can be conducted using expert judgment qualitatively based on a macro-thinking with regard to the above three factors, as detailed below.

5.2.2 Assessing CFP range

The CFP range of a product group is simply the difference between the highest CFP and lowest CFP of products in the group. The absolute value will be taken, i.e.: how much difference in kg CO_2e per GBP (Great Britain Pound) spent on the product.

A full scale formal CFP estimate based on LCA for those product groups is not necessary at this early stage. For the purpose of initial screening, a heuristic method is developed to get a quick and acceptable estimate. It has the following three steps.

Step 1: Search the existing literature for the CFP of a similar product. This similar product will serve as a base product for the estimation. It is ideal if the production of the base product requires the same materials and technologies with the product in question. Otherwise adjustments based on our best knowledge and understanding will be made.

Step 2: Find the GHG emissions (kg CO_2e) at each life cycle phase of the base product. They will serve as the baseline for estimating how much GHG are emitted during each life cycle phase of the product in question.

Step 3: Estimate the GHG emissions (kg CO_2e) at each life cycle stage of the product in question by adjusting the baseline according to the different attributes (weight, country of origin, etc.) of the product in question. Add the estimated GHG emissions (kg CO_2e) at each stage together to obtain the total CFP of the product. For example, for the group of TVs, the two products (32'' LCD-TV and 42'' PDP-TV) examined by Fraunhofer IZM, Öko-Institut, BIO Intelligence Service, Deutsche Umwelthilfe, PE Europe and CODDE(2007) are used as the base products (Step 1). From this source, the GHG emissions of the two base TVs at each life cycle stage are identified and listed in Table 4 (Step 2).

In Step 3, for the 6 TVs selected in the study we made some assumptions and concluded that based on those assumptions, their GHG emissions at the Production, Distribution and End of Life stages are about proportional their screen sizes. In the Use stage, the emissions are proportional to their energy consumptions. Suppose each TV has a 10 year life with 4 hours on and 20 hours stand by each day, their CFP estimates are given in Table 5.

"CFP range" is set as a quantitative criterion, and the worst and best values are set on the global scale applicable to all product groups. It is assumed, based on the CFP (in kg CO_2e per GBP) data at hand from literature, that the smallest possible range in a product group is 0 kg CO_2e per GBP, and the largest possible range is 6 kg CO_2e per GBP. The utilities for a few equidistant points in the range [0, 6] are shown in Table 6. The CFP ranges of the product group of TVs and Mobiles are presented in Tables 7 and 8 respectively.

5.2.3 Assessing consumer willingness and ability to shift

Many factors influence consumers' current preference of a certain product to another. These factors include healthiness, quality, convenience, price, etc. of the product, and financial status, technological proficiency, religious belief, etc. of the consumer. Now CFP becomes another factor to influence consumers' purchase decision.

Logically, if the CFPs of a group of products and the consumer current preferences for these products are positively correlated for reasons other than CFPs, i.e., the products preferred by consumers in the group "incidentally" have higher CFPs, their willingness and ability to shift to lower CFP products will be low; and if the CFPs and the consumers' current preferences are negatively correlated, their willingness and ability to shift to lower CFP products will be high.

Therefore, the correlation between the CFPs of the products and their relative consumer preferences is used as the indicator of the consumer willingness and ability to shift.

A proxy indicator of consumer preference of a product is the sales volume of this product. This is based on the assumption that product A sells better than product B if the consumers collectively prefer A to B. Therefore, consumer willingness and ability to shift can be indicated by the correlation between the CFPs of the products and their relative sales volume (kg for Food, liter for Drink, and pieces for EE (Electricals & Electronics) and Clothing) within one group of substitutable products.

"Consumer willingness and ability to shift" is set as a quantitative criterion. The utilities for a few equidistance points of the correlation within the range [-1, 1] are shown in Table 9.

Table 4 GHG Emissions of 32'' LCD & 42'' PDP TVs at Each Life Cycle Stages

Life Cycle Phase	Production			Distrib.	Use	End of Life			Total
	Material	Manuf.	Total			Disposal	Recycl	Total	
32'' LCD GHG (kg CO2e)	180	33	213	31	1072	32	-30	2	1318
42'' PDP GHG(kg CO2e)	292	331	623	17	2312	41	40	1	2952

Table 5 Estimates of GHG Emissions (kg CO_2e) of Selected TVs

Product	Production			Distrib.	Use	End of Life			Total
	Material	Manuf.	Total			Disposal	Recycl	Total	
A			245.46	35.72	942.06			2.30	1225.56
B			180.54	26.28	1201.94			1.70	1410.44
C			316.73	46.10	1110.98			2.97	1476.78
D			596.12	16.27	2323.95			0.96	2937.29
E			649.88	17.73	2300.05			1.04	2968.71
F			958.46	26.15	2718.24			1.54	3704.40

Table 6 Utility Functions for Value Points on "CFP Range"

CFP Range (kg CO2e per GBP)	Utility
6	1
4.5	0.75
3	0.5
1.5	0.25
0	0

Table 7 CFP Range of the Product Group of TVs

Product	CFP (kg CO2e/Item)	Price (£/Item)	CFP (CO2e/£)
A	1225.56	400	3.06
B	1410.44	450	3.13
C	1476.78	990	1.49
D	2937.29	800	3.67
E	2968.71	650	4.57
F	3704.40	1330	2.79
CFP Range			3.08

Table 8 CFP Range of the Product Group of Mobiles

Product	CFP (kg CO2e/Item)	Price (£/Item)	CFP (CO2e/£)
A	41.51	80	0.52
B	48.69	110	0.44
C	52.74	90	0.59
D	55.94	100	0.56
E	57.07	140	0.41
F	63.94	180	0.36
CFP Range			0.23

Table 9 Utility Functions for "Consumer Willingness and Ability to Shift (W&A)"

W&A indicator: Correlation between CFPs and sales volumes (kg for Food, liter for Drink, and pieces for EE and Clothing)	Consumer W&A	
Value points: correlation coefficient	Grades	Utility
-1	Very high	1
-0.5	High	0.75
0	Medium	0.5
0.5	Low	0.25
1	Very low	0

The correlations between the CFPs and sales volumes of the products in TVs and Mobiles groups are presented in Tables 10 and 11 respectively.

5.2.4 Assessing scale of product group

The scale of a product group is indicated by its total sales volume (£m). When assessing the opportunity, it is assumed that the larger the scale the better. Based on the group sales volume data provided by the retailer, we divide the sales volume broadly into five bands and associate each band with an assessment grade as listed in Table 12. Therefore the "Scale of product group" is set as a qualitative criterion. The utility functions for the assessment grades are shown in Table 12.

The distributed assessments of the group scale for the product groups of TVs and Mobiles are shown in Table 13.

5.2.5 Generating assessment outcomes of product shift related opportunity

The initial assessments of the three factors (or sub criteria)

Table 10 Correlation Between CFPs and Sales Volumes of the Products in the Group of TVs

Product	CFP (kg CO2e/Item)	Sales Volume (Item)
A	1225.56	3242
B	1410.44	6870
C	1476.78	90546
D	2937.29	5515
E	2968.71	2928
F	3704.40	1393
Correlation Coefficient	-0.237	

Table 11 Correlation between CFPs and Sales Volumes of the Products in the Group of Mobiles

Product	CFP (kg CO2e/Item)	Sales Volume (Item)
A	41.51	173355
B	48.69	32305
C	52.74	40040
D	55.94	6525
E	57.07	59605
F	63.94	3960
Correlation Coefficient	-0.426	

Table 12 Utility Functions and Definitions for Assessment Grades on Scale of Product Group

Assessment grades	Utility	Definitions for assessment grades Indicator: Sales Volume (SV) (£m)
Best	1	SV in [424.1,530]
Good	0.75	SV in [318.2,424.1]
Average	0.5	SV in [212.3,318.2]
Bad	0.25	SV in [106.4,212.3]
Worst	0	SV in [0.5,106.4]

Table 13 *Distributed Assessments of Scale of the Product Group*

	TVs	Mobiles
Scale of the product group	(Bad, 0.85) (Average,0.15)	(Worst,0.7) (Bad,0.3)

Table 14 Values/assessments of the sub-criteria under "Product shift related opportunity" for each product group

Product shift relates opportunity	TVs	Mobiles
CFP range (0.333)	3.08	0.23
Consumer W & A (0.333)	-0.237	-0.426
Scale of product group (0.333)	(Bad, 0.85) (Average,0.15)	(Worst,0.7) (Bad,0.3)

discussed above (and summarized in Table 14) can then be aggregated to generate a distribution (Figure 3) and an impact or utility score (see Figure 4), to indicate the ranking of the assessed product groups on "Product-shift related opportunity". The aggregation is carried out using the evidential reasoning algorithm (see Section 4) run behind the scene in the IDS software.

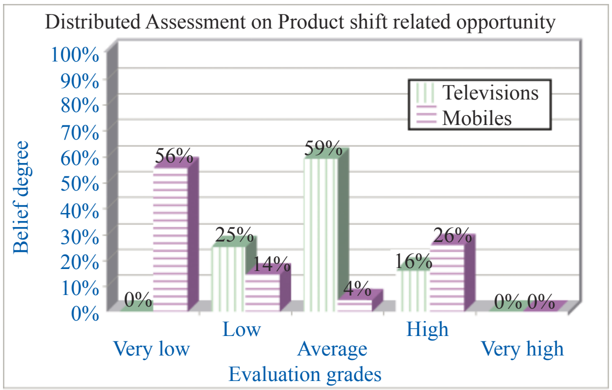

Figure 3. Distributions of product-shift related opportunity of labeling TVs and mobiles.

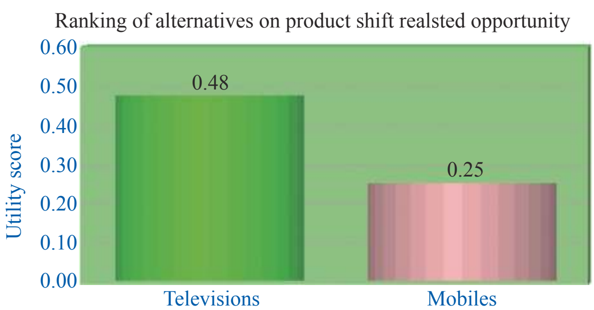

Figure 4. Utility scores of TVs and mobiles assessed against product-shift related opportunity.

5.2.6 Generating overall effect of labeling a product group

Following a similar process of assessing Product Shift Related Opportunity, the assessment of the two groups against other criteria can also be conducted. Further details are omitted here and available on request. Once the assessments are conducted against all criteria in the lowest level of the hierarchy, the assessment information is then aggregated in the same way as outlined in Section 5.2.5. For each product group, the aggregated outcomes include a distribution (see *Figure 5*) and a corresponding utility score (see *Figure 6*) on the top level criterion indicate the overall effects of labeling the product group. Similar distributions and utility scores on each of all other criteria in the hierarchy are also readily available for inspection in IDS. For example, the aggregated utility scores of the TVs group and the Mobiles group on the top, second, and third level criteria are given in Table 15.

Table 15 Utility scores of the TV Group and the Mobile Group

Opportunities and risks	TVs	Mobiles
Overall Effect (Average)	0.5037	0.3167
Carbon opportunity	0.5130	0.2432
Product shift related opportunity	0.4771	0.2507
Product improvement related opportunity	0.5667	0.3019
Non-carbon risk	0.4919	0.4374
Ethical risk	0.6518	0.4092
Commercial risk	0.3151	0.2753
Communication risk	0.5000	0.5500
Pollution risk	Unknown	Unknown
Resources risk	Unknown	Unknown

In Table 15 the utility scores of labeling program on "Pollution risk" and "Resources risk" are indicated as "*Unknown*" due to the limited information and resources we have for the analysis. Depending on whether the unknown information turns out to be in favor or in opposition of the assessed product group, its potential effect on the aggregated utility scores are bounded as indicated by the grey areas in Figure 6. From the figure, it is clear that the highest possible score of Mobile phones is still lower than the lowest score of TVs no matter what the unknown information is. Therefore if we are only interested in the relative ranking of the two, there is no need to collect further information about the impact of labeling the two groups on Pollution and Resource risks. The overall scores of 0.5037 and 0.3167 for TVs and Mobiles respectively correspond to the middle points of the corresponding grey areas and are used as an indicator for ranking.

In addition to the scores, IDS generates distributions which reveal the composition of negative, positive and unknown impacts of labeling a product group. For example, for the TVs and Mobiles product groups, the performance distributions are displayed in Figure 5. From each performance distribution, a utility range and a middle point can be calculated using equations (7) to (9). Compared with a score, a distribution contains richer information, supports a more transparent decision making process and decision makers are explicitly informed of the risks and opportunity of a decision.

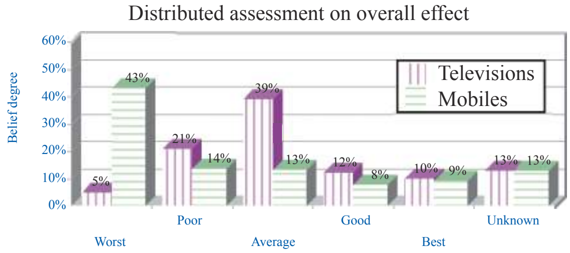

Figure 5. Distributions of overall effects of labeling TVs and mobiles.

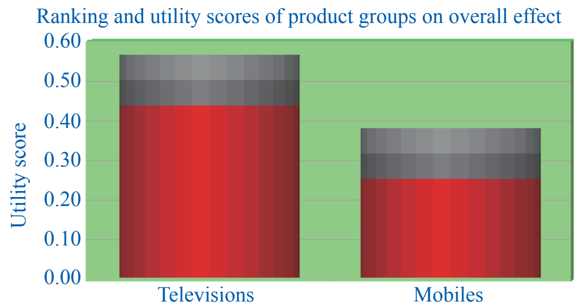

Figure 6. Overall effect utility scores of labeling TVs and Mobiles and ranking (the higher the scores, the higher the ranking).

5.2.7 Priority ranking

Generally the priority ranking of the product groups for carbon label consideration is based on their overall scores –the higher the scores, the higher the ranking. From Figure 6, TVs group has a higher utility score on Overall Effect, therefore higher priority should be given for further carbon labeling attention.

The high priority of the TVs group can be verified by further drilling down the criteria hierarchy. It can be seen from Table 15 that the TVs group has higher utility scores on Overall effect, Carbon opportunity, Non-carbon risks and almost all the third level criteria (except Communication risk) than the Mobiles group.

By following a similar process, among all the eight product groups analyzed in the study (as listed in Figure 2), the priority order is Meat, Cheese, Televisions, Fruit Juices, Mobile Phones, Wine, Women's Knickers and Women's Dresses.

5.2.8 Uncertainty analysis

Uncertainty analysis is a process of identifying the effect of one or various sources of uncertainty and unknowns on decision outcomes. In this study, uncertainties come from many sources. Lack of information in the assessment of Pollution Risk and Resource Risk, is one of them. Its effects on the priority scores are indicated by the grey areas on top of each bar in Figure 6. The grey areas will be reduced or diminished when more information becomes available and entered into the tool.

Another common source of uncertainty and also present in this study is the subjectivity of criterion weights. To analyze whether the uncertainty may affect the priority ranking, we may resort to the sensitivity analysis function of IDS as shown in Figure 7. The figure shows that no matter what weight is assigned to the Carbon Opportunity criterion (relative to the weight assigned to Non-Carbon Risks), TVs group always has a higher score than that of Mobile Phones group. This indicated the outcome is not sensitive to the changes of the weight. If the two lines cross at a point, then the weight will affect the ranking and the decision and its level needs to be carefully justified.

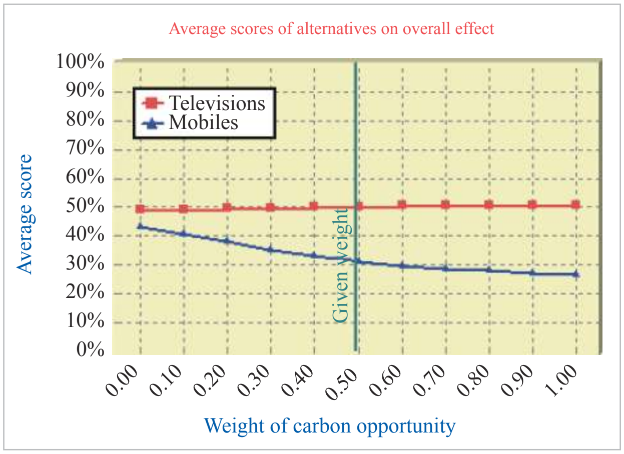

Figure 7. Sensitivity of ranking or utility scores to weight.

5.2.9 Trade-off analysis

Trade-off analysis is conducted in a two dimensional space, such as the Carbon Opportunity dimension and Non-carbon Risks dimension. Assessed product groups are positioned in the space according to their utility scores on each of the two dimensions. The picture (see Figure 8) helps decision makers to weigh the loss and gains of taking one course of action instead of the others.

In the picture, the top right corner of the green area representing an ideal outcome with high opportunity and low risks (see *Figure 8*). If the ideal outcome cannot be achieved, product groups positioned closer to the ideal point should have higher priority.

6 Conclusions

In this paper, a framework including assessment model methods is proposed to support the prioritization analysis of product groups for carbon labeling under various types of uncertainties. With the support of the IDS software, a tool is implemented and the process of its application is illustrated using examples. The outcomes of the analysis include the

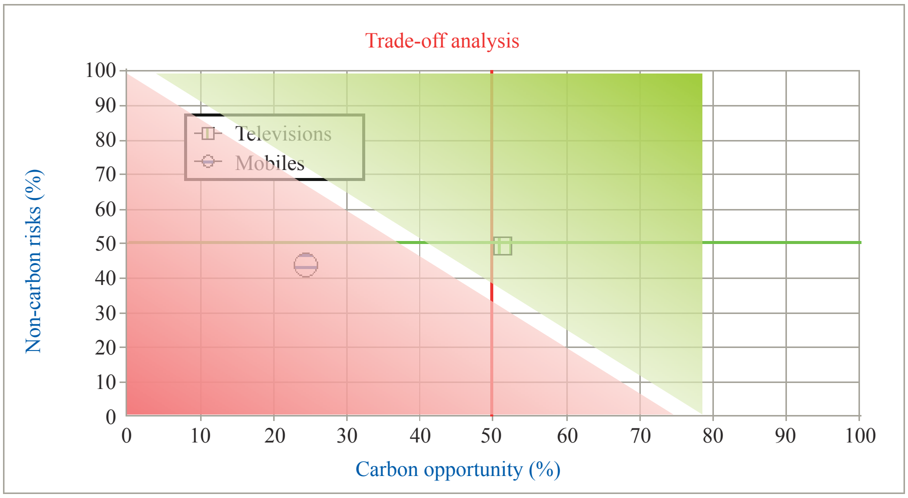

Figure 8. IDS graphical interface for trade-off analysis.

opportunity and risks, and the combined effects associated with the labeling of the eight selected product groups. Priority ranking of the eight groups based on the overall effects is given. The robustness of the results has also been examined through sensitivity analysis.

The assessment and the analysis of the eight product groups using the framework have led to the following key findings.

(1) The product groups in the Food category are associated with higher carbon reduction opportunity and more favorable overall effect than the product groups in other categories when it comes to carbon labeling. The Meat group, in particular, is associated with significantly higher opportunity and better overall effect than all the other groups.

(2) The product groups in the Clothing category are associated with the lowest carbon reduction opportunity and the most unfavorable overall effect. The Women's dress group, in particular, is associated with significantly worse overall effect than all the other groups.

(3) Within the electrical and electronic (EE) category, the TVs group is associated with higher carbon reduction opportunity and better overall effect than the Mobiles group.

(4) Super scale product groups with huge sales volume are always associated with high carbon opportunity and overall effect.

Based on the findings, it is recommended that high priority should be given to the Food category for carbon labeling attention. Other product groups which are not analyzed in the study but have huge sales volume, like the vegetable and bread groups should be considered and analyzed.

Due to the limited time and resources available to this study, the impact of product carbon labeling on pollution and resource risks is not assessed. Though analysis can still be carried out and in the case of TVs and Mobiles conclusion is reached without the need of further data collection.

In addition to the key findings, it can also be seen that the assessment model and process developed in the study requires relatively small amount of information and judgments, and can be applied to relatively large scale and systematic analysis of product prioritization for carbon labeling. Data can be collected through literature, retailer's product information and sales figures, statistical analysis, expert judgments and common sense rules. The framework is also general and can be useful to other retailers. If some data are difficult to obtain, using this framework, the effect of the missing data can be explicitly analyzed. The analysis can reveal whether the effect of the missing data is significant enough to affect the ranking of the options to be prioritized. If not, further data collection is unnecessary to reach a conclusion and time and efforts can be saved. However, when other product groups are added to the comparison, the ranking may not be conclusive.

Another issue is the assignment of criteria weights which is not systematically studied. Although sensitivity analysis can examine the effects of changing one criterion weight at a time, the current implemented tool cannot reveal the effects of simultaneous changes of all weights.

It should be noted that the accuracy of the carbon foot print estimates in this study is not high which is acceptable at this early stage of product screening but is not suitable for appearing on the label. A more accurate estimate based on a formal life cycle assessment should be conducted for each of those products identified as having high priority.

References

Beynon, M. (2005). Understanding local ignorance and non-specificity in the DS/AHP method of multi-criteria decision making. *European Journal of Operational Research,* 163(2), 403-417.

Bolwig, S., & Gibbon, P. (2009). *Emerging product carbon footprint standards and schemes and their possible trade impacts.* Report

Risø-R-1719(EN), Roskilde: Danish Technical University. Retrieved form
http://orbit.dtu.dk/fedora/objects/orbit:81457/datastreams/file_4013003/content

Belton, V., & Stewart,T. (2002). *Multiple Criteria Decision Analysis——An Integrated Approach*. Dordrecht□ Kluwer Academic Publishers.

Chin, K., Yang, J., Lam, J., & Guo, M. (2009). An evidential reasoning-interval based method for new product design assessment. *IEEE transactions on Engineering Management,* 56(1), 142-156.

Dyker, J., Fishburn, P., Steuer, R., Wallenius, J., & Zionts, S. (1992). Multiple criteria decision making, multiattribute utility theory: the next ten years. *Management Science*, 38(5), 645-654.

Fuentes, C., & Carlsson-Kanyama, A. (2006), *Environmental information in the food supply system. Swedish Defence Research Agency(FOI-R-1903-SE • ISSN 1650-1942).* Retireved form School of Business Economics and Law, Goteborg University,
http://www.fcrn.org.uk/sites/default/files/Environmental_information_in_the_food_supply_system.pdf

Fraunhofer IZM, Öko-Institut, BIO Intelligence Service, Deutsche Umwelthilfe, PE Europe and CODDE (2007), Ecodesign of EuP Products Preparatory Studies LOT 5: Consumer Electronics: TV. http://www.ecotelevision.org/finalised_documents.php.

Figueira, J., Greco, S., & Ehrgott, M. (2005). *Multiple Criteria Decision Analysis: State of the Art Surveys. Springer's International Series in Operations Research and Management Science*. New York: Springer Science Business Media.

French, S. (1988). *Decision theory - An Introduction to the Mathematics of Rationality*, Chichester: Ellis Horwood.

French, S., & Xu, D.(2005). Comparison study of multi-attribute decision analysis tools. *Journal of Multi-Criteria Decision Analysis*, 13(2-3), 65-80.

Goodwin, P., & Wright, G. (1999). *Decision Analysis for Management Judgment(2nd Edtion)*. New Jersey: Wiley.

Hwang, C., & Yoon, K. (1981). *Multiple Attribute Decision Making Methods and Applications*, Berlin: Springer-Verlag.

Jacobsen, G. (2011), The Al Gore effect: An Inconvenient Truth and voluntary carbon offsets. *Journal of Environmental Economics and Management*, 61(1), 67–78.

Kong, G., Xu, D., Body, R., Yang, J., Mackway-Jones, K., & Carley, S. (2012). A belief rule-based decision support system for clinical risk assessment of cardiac chest pain. *European Journal of Operational Research*, 219(3),564-573.

Keeney, R., & Raiffa, H. (1993). Decisions With Multiple Objectives(2nd ed). Cambridge: Cambridge University Press.

Liu, J., Yang, J., Ruan, D., Martinez, L., & Wang, J. (2008). Self-tuning of fuzzy belief rule bases for engineering system safety analysis. *Annals of Operations Research*, 163(1),143-168.

Maddulapalli, A., Yang, J., & Xu, D. (2012). Estimation, modelling and aggregation of missing survey data for prioritizing customer voices. *European Journal of Operational Research,* 220,762-776.

Shafer, G. (1976). Mathematical Theory of Evidence. Princeton: Princeton University Press.

Sii, H., Wang, J., Eleye-Datubo, A., Yang, J., & Liu, J.(2005). Safety assessment of FPSO turret-mooring system using approximate reasoning and evidential reasoning. *Marine Technology and SNAME News*, 42(2), 88-102.

Savage, S. (2009) The Flaw of Averages (2009), New Jersey: Wiley.

Stewart, T. (1992). A critical survey on the status of multiple criteria decision making theory and practice. *OMEGA International Journal of Management Science*, 20(5-6), 569-586.

The Carbon Trust. (2008). *Product carbon footprinting: the new business opportunity Experience from leading companies*. Retrieved form carbontrust website: http://www.carbontrust.com/media/84928/ctc744-product-carbon-footprinting-the-new-business-opportunity.pdf.

Teisl, M., Roe, B., & Hicks, R. (2002). Can eco-labels tune a market? Evidence from dolphin-safe labeling. *Journal of Environmental Economics and Management,* 43, 339-359.

Udo de Haes, H., Finnveden, G., Goedkoop, M., Hauschild, M., Hertwich, E., & Jolliet, P.O. (2002). *Life-Cycle Impact Assessment: Striving Towards Best Practice*. Florida: SETAC press.

Wang, J., Sii, H., Pillay, A., Yang, J., Kim, S., & Saajedi, A.(2004). Use of advances in technology in marine risk assessment. *Risk Analysis*, 24(4), 1011-1033.

Wang, Y., Yang, J., & Xu, D. (2006). Environmental impact assessment using the evidential reasoning approach. *European Journal of Operational Research*, 174(3), 1885-1913.

Xu, D., McCarthy, G., & Yang, J. (2006). Intelligent decision system and its application in business innovation self-assessment. *Decision Support Systems*, 42, 664-673.

Xu, D., Yang, J., Carlé, B., Hardeman, F., & Ruan, D. (2008). Application of intelligent decision system to nuclear waste depository option analysis. *International Journal of Nuclear Governance, Economy and Ecology,* 2(2), 146-165.

Xu, D., Yang, J., & Wang, Y. (2006). The ER approach for multi-attribute decision analysis under interval uncertainties. *European Journal of Operational Research,* 174(3),1914-1943.

Yang, J. (2001). Rule and utility based evidential reasoning approach for multiple attribute decision analysis under uncertainty. *European Journal of Operational Research*, 131, 31-61.

Yang, J., & Singh, M. (1994). An evidential reasoning approach for multiple attribute decision making with uncertainty. *IEEE Transactions on Systems, Man, and Cybernetics*, 24(1), 1-18.

Yang, J., & Xu, D. (2002). On the evidential reasoning algorithm for multiattribute decision analysis under uncertainty. *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans,* 32(3), 289-304.

Yang, J., & Xu, D. (2011). Introduction to the ER rule for evidence combination in Integrated Uncertainty in Knowledge Modelling and decision Making, *Springer*, 7027, 7-15,

Yang, J., Xu, D., Xie, X., & Maddulapalli, A., (2011). Multicriteria evidential reasoning decision modeling and analysis – prioritizing voices of customer. *Journal of the Operational Research Society*, 62,1638-1654.