A comparison of methods for calculating the carbon footprint of a product

A. Barnett[1]*, R. W. Barraclough[2], V. Becerra[3], S. Nasuto[3]

[1] Technologies for Sustainable Built Environments Centre, University of Reading, United Kingdom
[2] @UK PLC, Aldermaston, United Kingdom
[3] School of Systems Engineering, The University of Reading, United Kingdom

* corresponding author: a.j.barnett@pgr.reading.ac.uk

Abstract

Carbon footprinting is a way of estimating greenhouse gas emissions caused by an organization, event, product or person. Methods for carbon footprinting are relatively new, and there are two emerging approaches: the process model and the input-output model. The carbon trust provided data from 365 products, that is analysed in this paper using two such methods. One method is an input-output model developed by the Centre for Sustainable Accounting and implemented by @UK PLC. The other is a process model analysis conducted by PAS2050, a standard which lays out how to assess greenhouse gas emissions of a products life cycle. In addition to direct comparison of the results we use spectral clustering to examine relative similarities in the outputs of the two techniques and try to determine areas where the ratio between the carbon footprint of products is similar across the two techniques. The outcome of the direct comparison shows that the resulting value of the two footprinting techniques have no significant direct linear relationship. The results of the spectral clustering show that, based on the selected similarity graph, there are no distinct clusters, and hence no meaningful relationship between the relative differences of products, across footprinting techniques.

KEYWORDS

Product carbon footprint, PAS2050, environmental impact, spectral clustering.

1 INTRODUCTION

The scientific consensus on climate change is that: 'the climate is changing and that these changes are in large part caused by human activities' (America's Climate Choices et al., 2010). The main contributing factor to climate change is the level of *greenhouse gasses* – such as carbon dioxide – in the atmosphere and, in turn, the rate at which human activity is releasing further greenhouse gasses into the atmosphere (America's Climate Choices et al., 2010).

Carbon footprinting is the process of quantifying the greenhouse gas emissions associated with human activity. Usually a carbon footprint is associated with the production of a product, the delivery of a service or, more generally, the occurrence of some process. The Kyoto protocol (United Nations, 1998) defines the six main greenhouse gasses. By weighting each of the gasses according to the damage it causes to the environment the protocol allows carbon footprint to be quantified in terms of a single measure: *kilogrammes of carbon dioxide equivalent*.

Carbon footprinting is a new discipline; there are different definitions of carbon footprint (Wiedmann and Minx, 2008), and different approaches to quantifying it. In this paper we consider carbon footprint to mean: *the greenhouse gas emissions of a process or the life cycle of a product measured in kilograms of carbon dioxide equivalent*. We shall compare implementations of the two prevailing methods for quantifying carbon footprint: the *process* model, and the *input-output* model.

All carbon footprints have a *scope* – though this may be implicit in the method by which the carbon footprint is quantified. Since no process is completely isolated, boundaries must to be set regarding what to include in the carbon footprint. For example a product carbon footprint could take into account the acquisition of raw materials and the carbon footprint associated with disposal of the product, or purely the manufacturing of the product from the raw materials. It has been shown that the definition of scope can result in significant differences in the carbon footprint assigned to a product or process (Lenzen and Dey, 2000).

A fundamental problem in quantifying carbon footprint is the availability of data upon which to base any calculations, and the amount of and detail required in such data. For example, the country in which a product is manufactured affects how far it must travel to a UK based consumer. This in turn affects its carbon footprint.

Carbon footprints are difficult to quantify, and since there is no known 'right answer' it is even more difficult to assess the accuracy of the methods for quantifying carbon footprint. In this paper we are comparing two methods for quantifying carbon footprint. Rather than saying that one is more accurate than the other – which is necessarily impossible – we wish to determine whether there is any systematic relationship between the carbon footprints quantified by the two methods. For example: does one method consistently under-estimate the carbon footprint relative to the other?

2 CARBON FOOTPRINT MODELS

Examples of the two different product carbon footprinting models analysed in this paper are the *process model* and the *input-output model*.

2.1 Process model carbon footprints

The PAS2050 method (British Standards Institute, 2008) is an example of a process model. A process model is a *bottom up approach*. The model takes into account all processes in the product life cycle, from production to disposal of the product and because of this, logically it is assumed to be the most accurate model. An example of a PAS2050 assessment of a product has been performed on croissants by the Carbon trust (2008).

The process model can be more accurate than the input output model on a product specific level, for example if the manufacturer of the product has utilised low carbon sources, the footprint generated will reflect that.

The method requires detailed information on the entire life cycle of the product and so is very expensive, in terms of time and computation. A major issue with the process model is that the data required is often not available causing reductions in accuracy. This for example could possibly be due to a supplier not wanting to divulge information on their production processes. It is a manual process and can take days per product so is impractical for large scale use.

The process model can be more accurate than the input output model on a product specific level, for example if the manufacturer of the product has utilised low carbon sources, the footprint generated will reflect that.

2.2 Input-output model carbon footprints

The method used by @UK PLC is based on the input-output model developed by the Centre for Sustainable Accounting (Wiedmann et al., 2007). This model is a *top down approach*. The model uses carbon intensities, measured in kilograms of carbon dioxide per pound spent, to assign footprint to a product based on the price of the product. @UK PLC developed an auto classifier system (Roberts, 2010) to assign the product to a category, then the spend of the product is multiplied by the carbon intensity of that category, producing a carbon footprint.

The system is fully automated and so costs very little in terms of both computation and time. The system also requires little information of the production of the product itself, only requiring a price, a description and a unit of measure. Because of this the method is very quick and can handle large catalogues of data in a short period of time. The method is used currently in a product called *GreenInsight* (@UK-plc, 2012), which is commercially available.

The main issues with the method is that, because the model is based on sector averages, it cannot handle any product specific data, for example low carbon sourcing.

3 THE DATA SET

The data set used was 1551 products, footprinted using the PAS2050 carbon footprinting method by the Carbon Trust (source no longer available). Of these products only 365 produced comparable data when a carbon footprint was generated using the @UK PLC method. The incomparable data was due to a lack of pricing information and incomparable units of measure. For example, the PAS2050 footprint of shower gel was given per shower. Shower gel is however sold per bottle and so without further information detailing how many showers a bottle of shower gel can be used for, the comparison cannot be completed.

The categories that the usable data covered were:

- Bags and sacks (27 products)
- Bakery products (5 products)
- Books and magazines (1 products)

- Crisps and snacks (2 products)
- Dairy products (11 products)
- Fruit and vegetable juices (27 products)
- Jams and spreads (1 products)
- Non alcoholic beverages (6 products)
- Pasta, cooked and stuffed (58 products)
- Soft drinks and mineral water (13 products)
- Tableware (3 products)
- Vegetables (22 products)
- Wearing apparel (except fur) (189 products)

For comparison the products were assigned to the relevant categories using manual classification (to avoid misclassification and ensure the comparison was as accurate as possible). This is only possible for data sets of limited size.

4 SPECTRAL CLUSTERING

Spectral clustering is a family of methods for determining similarities in data. Several variants are laid out by Luxburg (2007). Clustering is used to determine relationships between data points in multidimensional data. If points are clustered, it means that they share common properties. Figure 1a from Ng et al. (2002) shows how spectral clustering can be used to discriminate between different letters.

The clustering algorithm used in this paper is given by Luxburg (2007) and laid out below.

Let $P = \{p_1, p_2, \ldots, p_n\}$ be products and suppose that p_i has associated with it two carbon footprints: u_i assigned by @UK PLC and t_i assigned by the Carbon Trust. Define matrices

$$A_u = (a^u_{ij}) \text{ where } a^u_{ij} = \begin{cases} \frac{u_i}{u_j} & \text{if } i \leqslant j \\ \frac{u_j}{u_i} & \text{otherwise} \end{cases} \tag{1}$$

$$\text{and } A_t = (a^t_{ij}) \text{ where } a^t_{ij} = \begin{cases} \frac{t_i}{t_j} & \text{if } i \leqslant j \\ \frac{t_j}{t_i} & \text{otherwise.} \end{cases} \tag{2}$$

Note that A_u and A_t are both symmetric. Next define

$$A = (a_{ij}) \text{ where } a_{ij} = |a^u_{ij} - a^t_{ij}| \tag{3}$$

which is again symmetric. The element a_{ij} of A may be interpreted as the 'distance' between the carbon footprints of p_i and p_j as computed by the two different systems.

Define the diagonal matrix D by

$$D = (d_{ij}) \text{ where } d_{ii} = \sum_{j=1}^{n} a_{ij} \tag{4}$$

and $d_{ij} = 0$ when $i \neq j$. The Laplacian matrix used by the spectral clustering algorithm is $L = D - A$. The next step is to compute the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of L (including multiplicities) and their associated eigenvectors, $v_1, v_2, \dots, v_n$ say. Since L is positive semi-definite, all of the eigenvalues are real and non-negative. Therefore they can be ordered so that so that $\lambda_j \leqslant \lambda_{j+1}$ $\forall j\ 1 \leqslant j < n$. Let V be the $n \times k$ matrix whose columns are the first k eigenvectors of L, and let w_i be the ith row of this matrix ($1 \leqslant i \leqslant n$).

The vectors are then clustered $w_i \in \mathbb{R}^l$ using the k-means algorithm to obtain clusters $C_1, C_2, \dots, C_k$ where $C_j \subseteq \{w_1, w_2, \dots, w_n\}$ and $C_i \cap C_j = \emptyset\ \forall i, j$.

Finally, the original data points are now clustered as:

$$\bar{C}_r = \{p_i \mid 1 \leqslant i \leqslant n \text{ and } w_i \in C_r\}. \tag{5}$$

The clustering defined above compares the quotients of carbon footprints assigned by the two methods, taken across all pairs of products. This will discover whether the relative difference in the carbon footprint of two products is consistent between the two methods – a linear relationship – or whether some subsets (clusters) of products support some similar relationship. This is in contrast to simply applying the k-means algorithm to the data points $(u_i, t_i) \in \mathbb{R}^2$ since for example linear relationship does not exhibit clusters here.

5 RESULTS

The direct comparison results given in figure 1 shows very little correlation between the footprint generated from the two methods. The line of best fit is provided using linear regression (Weisberg, 2005). Ideally the points would conform to this line. This would mean that a linear relationship exists, *i.e.,* all the footprints generated by one method were a multiple of those generated using the other method. This value would give a strong indication to a link between the two methods.

When the sum of squared errors are compared to the model errors the unexplained variance, R^2, is 0.130. The linear model is accurate if the value of R^2 is close to 1. This therefore shows the linear model to be an inaccurate representation of the data.

Spectral clustering was applied using two clusters. Thus the matrix V from section 4 is 365×2 and we can visualise easily its rows $w_i \in \mathbb{R}^2$, which are plotted in figure 2. The different colours represent different clusters and the *x's* represent the centroids of the clusters.

The graph shows that all but one of the points are assigned to one cluster and the final point is assigned to the other. This shows that the data is not partitioned in a way detectable by this technique. The graph does not give any information on the structure within the cluster as it is designed only to ascertain the existence of clusters within data.

Spectral clustering is mostly applied to data of dimension greater than 2, as applying the technique to two dimensional data will provide an equal level of information to visual inspection. The data provided, however, was two dimensional, therefore a two dimensional analysis was performed.

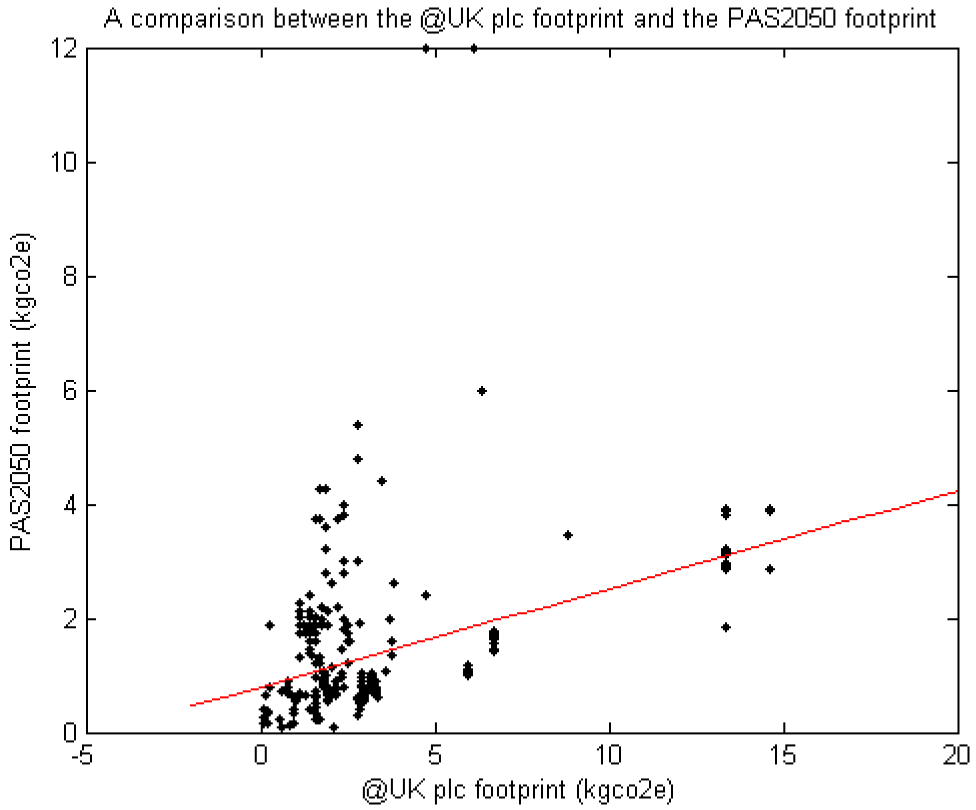

Figure 1: Direct comparison between @UK PLC footprint and PAS2050 footprint.

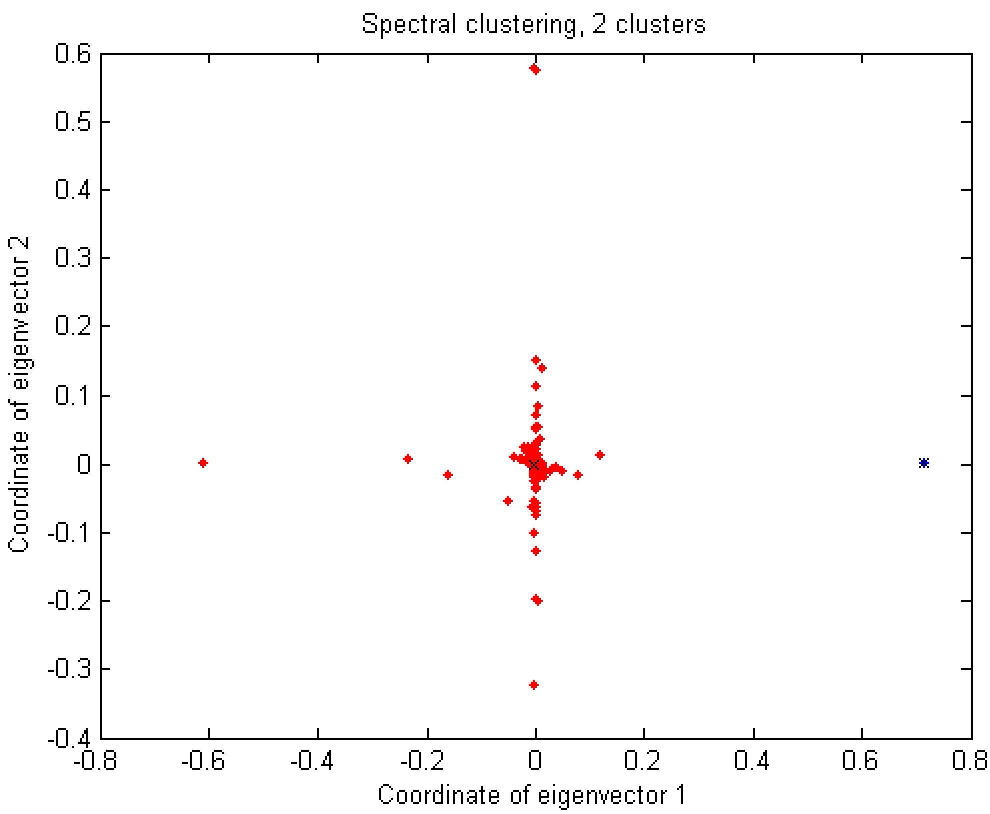

Figure 2: Spectral clustering results using the first 2 eigenvectors, with 2 clusters.

6 CONCLUSION

From the analysis techniques applied to the data in this paper, no significant correlation between the two carbon footprinting techniques can be determined. One potential reason for the difference was that the clothing within the data set—which makes up approximately half of the data set—was from a low carbon supplier. The PAS2050 method—which takes significantly more data into account—will therefore calculate a lower carbon footprint. This is because it would include the low carbon measures taken by the supplier. The @UK PLC method however will only classify it as clothing and— because the price is similar than that of standard manufacturers—the footprint will not take into account the low carbon measures that the supplier has utilised.

Another potential issue is the small volume of data. The data set is relatively small and cannot accurately represent the millions of products that exist. Because of this it is not possible to draw an overall comparison between the two methods, but it is possible to draw preliminary conclusions based on limited results.

The volume of data that can be used is determined by the number of carbon footprints that can be produced using the PAS2050 method. Because the PAS2050 method takes a large amount of time and computation to complete each carbon footprint, only a small number of carbon footprints can be produced using this technique in a feasible time frame.

The direct comparison technique attempted to find a linear relationship between the carbon footprints obtained from the two methods. The conclusion can however be drawn that no meaningful linear correlation between the two data sets exists. This result is emphasised by the results of the linear regression, as the unexplained variance, R^2, of 0.13 is relatively low.

The result of the spectral clustering are highly dependent on the similarity graph used. It is possible that using a different similarity graph will determine a different relationship between the two carbon footprinting techniques. The similarity graph was chosen to determine if a relationship between two products carbon footprint was consistent independent of footprinting technique used. The fact that all but one of the points were assigned to one cluster, however, indicates that no such relationship exists.

Any conclusions drawn from this paper cannot be assumed to be characteristic of product carbon footprinting in general, but can give an indication as to potential relationships, due to the limited size of the data set employed.

This paper focussed on comparing the results of the two carbon footprinting methods and not the methods themselves. There is however a significant difference in the techniques in terms of both time, computation and levels of data required to perform the analysis. The footprints generated using the PAS2050 method were provided from a repository, meaning it is only possible to speculate on the method used. However it is shown that process based carbon footprinting, on which the PAS2050 method is based, is largely expensive in terms of time, computation and quantity of data. It would almost certainly take longer and more computation than the @UK PLC method, which was completed within a matter of minutes.

7 FUTURE WORK

In the future a more detailed comparison of the results of the two methods is required to determine definitively if there is a functional relationship between them. The direct comparison has proven an absence of a strong linear relationship between the two techniques, therefore one potential technique to apply would be to attempt to find a non-linear relationship between the results of the two methods. This requires a more complex and detailed analysis.

The more detailed analysis will potentially include other variants on the similarity graph used in the spectral clustering. This will allow for different relationships to be tested, *i.e.,* if the difference between product footprint is consistent across the different carbon footprinting methods.

Additionally, further analysis will require a more comprehensive volume of data, as any conclusions drawn from a small dataset cannot be considered representative of the entire data range.

REFERENCES

America's Climate Choices, Panel on Advancing the Science of Climate Change and National Research Council (2010). *Advancing the Science of Climate Change*, The National Academies Press.
URL: *http://www.nap.edu/openbook.php?record_id=12782*

British Standards Institute (2008). *Publicly Available Specification PAS 2050: 2008 Specification for the assessment of the life cycle greenhouse gas emissions of goods and services*, BSI British Standards Institution.
URL: *www.bsigroup.com*

Carbon Trust, Department of Environment food and Rural affairs DEFRA and British standards Institute (2008). *Guide to PAS 2050 How to assess the carbon footprint of goods and services*, BSI Standards.

Lenzen, M. and Dey, C. (2000). Truncation error in embodied energy analyses of basic iron and steel products, *Energy* **25**(6): 577–585.

Luxburg, U. (2007). A tutorial on spectral clustering, *Statistics and Computing* **17**(4): 395–416.
URL: *http://www.springerlink.com/index/10.1007/s11222-007-9033-z*

Ng, A., Jordan, M. and Weiss, Y. (2002). On spectral clustering: analysis and an algorithm, *Advances in Neural Information Processing Systems* **14**: 849 – 856.

Roberts, P. J. (2010). *Automatic Product Classification*, PhD thesis, University of Reading.

@UK-plc (2012). @UK PLC GreenInsight.
URL: *http://www.green-insight.com/*

United Nations (1998). Kyoto Protocol To The United Nations Framework Kyoto Protocol To The United Nations Framework, *Technical report*.

Weisberg, S. (2005). *Applied Linear Regression*, third edn, Wiley-Interscience.

Wiedmann, T. and Minx, J. (2008). A Definition of Carbon Footprint, *in* C. Pertsova (ed.), *Ecological Economics Research Trends*, Nova Science Publishers, chapter 1, pp. 1–11.
URL: *https://www.novapublishers.com/catalog/product_info.php?products_id=5999*

Wiedmann, T., Wood, R., Lenzen, M. and Harris, R. (2007). Application of a novel matrix balancing approach to the estimation of UK input-output tables, *The International Input-Output Association conference*, Istanbul, pp. 1–46.
URL: *http://www.io2007.itu.edu.tr/*