Scientific Journal
Warsaw University of Life Sciences – SGGW

PROBLEMS OF WORLD AGRICULTURE

Volume 12 (XXVII)
Number 3

Warsaw University of Life Sciences Press
Warsaw 2012

Cecília Szigeti[1]
Anita Borzán[2]
Szent István University
Gödöllő, Hungary

Extreme outliers in the database for calculation of ecological footprint; the problems of grazing land footprint as well as the fishing ground footprint calculation

Abstract. The appreciation of ecological footprint has been increasing since the study by Stiglitz, Sen and Fitoussi' [Stiglitz et al. 2009]. At the same time, owing to the methodological and standardization problems as well as the shortcomings of data collection, its accuracy can be questioned. In our study, we were looking for countries with a significantly differing from the world average composition of the ecological footprint, with the help of cluster analyses and data from the database used for calculating the ecological footprint index by the Global Footprint Network. Comparing data from two years, we were trying to find answer to the question if the outlier data can trace back to professional errors or data collection problems. Basing on our studies, we can determine a two members group with an outstanding grazing land footprint (Mongolia, Uruguay), which can be considered as outliers according to every examination method. The formation of a stable group characterized by a big fishing ground footprint in 2010 can trace back to an inconsistency in the database, which is proved by the example of Gambia and Norway. In our opinion, a control of outliers is necessary for proper calculation of the EF index every year.

Key words: ecological footprint, grazing land footprint, fishing ground footprint, hierarchical cluster, Gambia.

Introduction

The Global Footprint Network (GFN)[3] calculates the ecological footprint (EF) for countries and the whole world. The national results as well as the global trends can be downloaded by land use categories from the website of the institution. The GFN prepares guidelines and information for the calculation [Kitzes et al. 2008; Ewing et al. 2010].

One of the most significant criticisms of the EF index territorial application is that the borders of countries have been established basing on geopolitical and cultural aspects. For this reason these by no means have any environmental meaning because they usually divide connected ecosystems. In this aspect, the EF calculation for territories within their natural borders is applicable for straighter conclusion. At the same time, nations are the largest decision-making bodies, so an environmental intervention can be made in the first place in this frame. For this reason, one of the suggestions by the spatial calculation critics is that the index should not be used in spatial instead of temporal analyses: 'The per capita EF is

[1] PhD, e-mail: szigetic@sze.hu, address: H-9026, Győr , Egyetem str. 1.
[2] PhD, e-mail: borzan.anita@gk.szie.hu, address: H-5600, Békéscsaba, Bajza str. 33.
[3] You can find detailed information about the institution and co-operation possibilities as well as the publications related to GFN in the web [Global…2010, 2011].

neither very informative about the spatial distribution of the impacts nor the causes of environmental pressure' [van den Bergh & Verbruggen 1999].

We are examining in our study the extreme outlier data included in the GFN database for 2010 and 2011. We are looking for those countries whose ecological footprint structure differs significantly from the average one. Comparing the data of the two years, we are trying to find answer for the question if the outlier data can trace back to professional errors or data collecting problems.

Material and method

The GFN database gives the ecological footprint for certain countries between 1961 and 2008 (Table 1). The increase of EF went together with the transformation of its structure, which means that the carbon footprint increased fivefold and the carbon dioxide emission is now responsible for more than half of the EF. Behind the seemingly unambiguous global tendency, significant individual national differences can be realized.

Table 1. Structure and size of global ecological footprint for between 1961 and 2008, global hectare (gha) / person

Ecological footprint and components	Year										
	1961	1965	1970	1975	1980	1985	1990	1995	2000	2005	2008
EF	2.4	2.5	2.8	2.8	2.8	2.6	2.7	2.6	2.5	2.6	2.7
cropland footprint (cr)	1.1	1.1	1.0	0.9	0.8	0.8	0.7	0.7	0.6	0.6	0.6
grazing land footprint (gr)	0.4	0.4	0.3	0.3	0.3	0.2	0.2	0.2	0.2	0.2	0.2
forest footprint (fo)	0.4	0.4	0.4	0.4	0.4	0.3	0.3	0.3	0.3	0.3	0.3
fishing ground footprint (fi)	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
carbon footprint (ca)	0.3	0.5	0.9	1.0	1.1	1.1	1.2	1.2	1.2	1.4	1.5
built-up land footprint (bu)	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1

Source: GFN database of 2011 [Global... 2011].

We conducted our analyses with the application of IBM SPSS20 programme package and we relied on the database analyses manual by Sajtos and Mitev [2007] for selecting the methods and assessment of the results. A ranking was performed with the help of cluster analysis.

We applied GFN database for 2010 and 2011 for our research [Global... 2010 & 2011]. We used a graphic method, a boxplot diagram for mapping the outliers. Then in the second part of our research we investigated if a linear relationship was realized between pairs from among six components of the EF. We indicated values of Pearson's correlation coefficients in form of a correlation matrix (Table 2). Since the cluster analysis is very sensitive to the appearance of outliers, we checked the outstanding data with the nearest neighbour method before every examination and excluded these values from the examination. From the perspective of assessment of results, it is important that we did not exclude the outstanding values of certain data, but only those which would have created a single group. The data used in the study are measured on the same metrical scale, for this reason we used not standardized data. In co-operation with two independent variables, trio as well as five variables we excluded the cropland footprint because of the strong multi-collinear method. We conducted hierarchical cluster analyses with analyses of variance: with 'Nearest neighbour' and Ward's methods. Clustering was performed, in case it was necessary, with K-means cluster , and not with hierarchical method, which was followed by summarising the results with analysis of contingency tables.

The aim of the cluster analyses in the first place was not to limit the country sets to countries with similar characteristics, but to identify the outstanding values as well as the outliers.

Findings

Based on the statistical data of 150 countries, Table 2 shows the linear correlation coefficients between the ecological footprint components. Pearson's correlation coefficients marked by bold typing show significant relationship between certain components of ecological footprint, while others are independent. Since strong relationship can not be noticed anywhere, in principle there is nothing to prevent us from withdrawing all of the variables from the cluster analyses. Table 2 shows results for the 2010 database, however, an investigation of the 2011 one revealed similar results.

Table 2. Matrix of Pearson correlation coefficients

EF component[4]	cr	gr	fo	fi	ca	bu
cropland footprint (cr)	x	x	X	x	x	x
grazing land footprint (gr)	-0.23	x	X	x	x	x
forest footprint (fo)	**0.334**	0.023	X	x	x	x
fishing ground footprint (fi)	**0.273**	-0.101	0.214	x	x	x
carbon footprint (ca)	**0.641**	0.008	**0.277**	**0.231**	x	x
built-up land footprint (bu)	**0.601**	-0.008	**0.293**	0.114	**0.352**	x

Source: own calculation based on GFN 2010 database [Global... 2010].

We conducted the first cluster analysis in our research with five variables, however we did it with three variables in our second study.

In our first examination, owing to stronger than mid-range relationship between cr and ca as well as cr and bu variables, we conducted the examination with 5 variables excluding cr. The outliers revealed with the nearest neighbour method, with regard to the data from 2010, Mongolia; Uruguay; Australia; Qatar and United Arab Emirates. We received similar

[4] Table 1 consists of the meanings of the abbreviations in the chart.

results with the use of database for 2011. Results obtained with similar methods are also similar, however, Kuwait appeared among the outliers instead of Australia. We repeated the cluster analyses with three, four and five cluster solutions without outlier values, nevertheless none of them provided appropriate results, the grouping of countries can not be performed clearly based on the examined variables.

Based on independent variables, trios (gr, fi, fo and fi, gr, bu) in pairs can be separated. Continuing our examination in order to eliminate the deviations because of correlation (Table 2), we repeated the analyses with two variables groups as well.

In our second study, we performed cluster analyses with the use of gr, fi, and fo variables and we excluded Mongolia and Uruguay outliers discovered by the nearest neighbour method. Regarding the method of Ward, used in the first case (the examination was done with 3 up to 7 clusters), we did not receive any appreciable results. Neither the nearest neighbour method analyses nor the K-means hierarchical cluster analyses did not result in satisfying solution.

We continued the examination with the other trio of independent variables (fi, gr, bu). After excluding the outliers (Mongolia, Uruguay), we came, basing on the database of 2010, to similar conclusion as when using the hierarchical cluster analysis as well as Ward's method. The best solution seemed to be dividing the countries into five clusters. According to two methods, three clusters were completely the same, in which a three members group was established, including Gambia, Mauritius and Norway. This group, which is considered to be stable, has not been established in 2011.

Results

Among the outliers revealed during the analyses, in all three cases Mongolia and Uruguay can be found in both examined years. We were looking for the reasons for this, while examining the composition of the ecological footprint in these two countries.

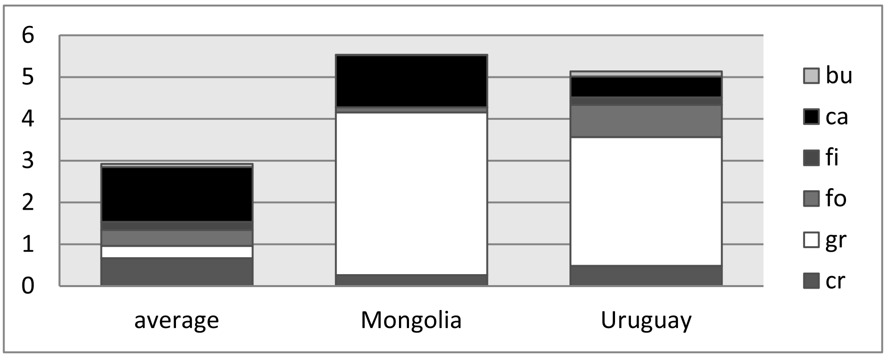

Fig. 1. Composition of the ecological footprint in selected countries, gha/person

Source: [Global... 2010].

As displayed in Figure 1, the most significant component of the average ecological footprint[5] is the carbon footprint (marked in black colour in the figure) and the grazing land footprint (marked in white in the figure) reaches 10% of the whole footprint. On the contrary, the grazing land footprint represents 70% in the ecological footprint of Mongolia and 60% in that of Uruguay, which are considered to be extreme outliers. Their grazing land footprint means 10 times more than the world average and this also means to be an extreme outlier (Figure 1).

In the boxplot diagram (Figure 2) the rectangle shows the distance between the top and the bottom quartile; the middle horizontal line is the median. The length of the vertical line is one and a half bigger than the extent of quartile. The outliers are the data that are out of the space between the extreme quartiles. If the data can be found outside of three times the extent between these quartiles, we call it extreme outliers and we sign it with *symbol. Figure 2 demonstrates excellently that the extreme outliers and the average values can be different from each other so significantly.

Figure 2 and Figure 4 in our study were created by the SPSS20 program. The result is the same when using the database of 201, which confirms that the two countries with their well-defined ecological footprint structure create a well distinguished separate group.

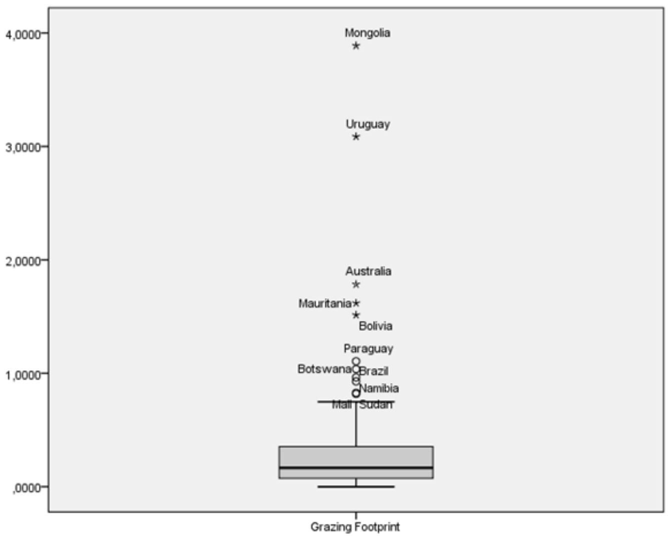

Fig. 2: Boxplot diagram of grazing-land footprint

Source: GFN 2010 database [Global... 2010].

Among the results of the cluster analyses conducted using data of 2010, the group with invariant and homogenous characteristics is interesting because of the differences between its members. The reason for this lies in the composition of their ecological footprint since

[5] There are some small differences between the arithmetic average of the national data shown here, in other words the ‘average ecological footprint’, and the composition and size of the global ecological footprint outlined in Table 1 due to methodological reasons.

we can realize that Gambia, Mauritius and Norway, with remarkably different geographical, cultural and economic characteristics, have one common feature, which is that their fishing ground footprint represents 38% up to 58% of their total footprint (Figure 3).

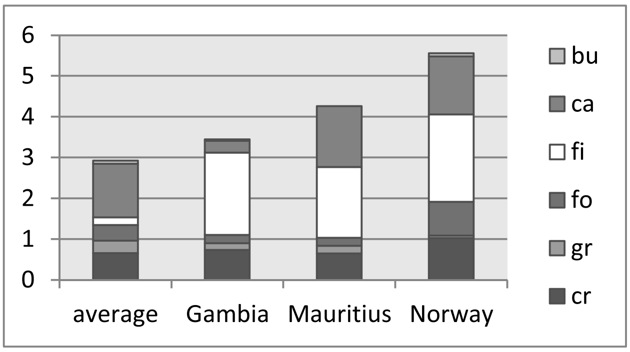

Fig. 3. Composition of the ecological footprint in selected countries, gha/person

Source: GFN database of 2010 [Global... 2010].

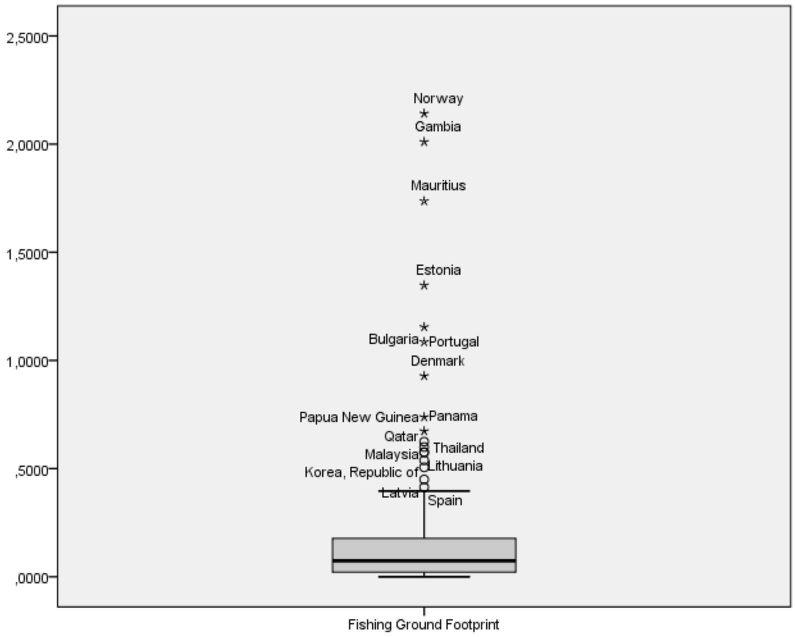

Fig. 4: Boxplot diagram of fishing-ground footprint

Source: own calculation based on the GFN database of 2010 [Global... 2010], SPSS20 output.

Figure 4 and Figure 2 can be interpreted similarly: they demonstrate the differences between the extreme outliers and the average data.

This prominently high fishing ground footprint values are 9 to 11 times more than the world average (Figure 4). When examining the database of 2011, Gambia can not be discovered among the extreme outliers, even its fishing ground footprint is the same as the world average of 0.1 gha/person, according to this database.

Conclusions

The appreciation of the ecological footprint indicator considerably varies in different application areas. While it is said to be the best indicator of 'unsustainability' on global level [Stiglitz et al. 2009], its spatial application is criticized from many angles [van den Bergh & Verbruggen 1999; McDonald & Patterson 2004]. However, the mentioned studies have not examined the whole database of the developing Global Footprint Network indicator, so the critical statements as well as the reservations have been conducted without screening the outliers. According to our study, the majority of the countries in the world (112 out of 150) can be described by an average EF structure. In other words, we can come to the conclusion about the ecological footprint composition of certain countries basing on the average ecological footprint composition. However, there are some well distinguished small groups, which have an EF composition that significantly differs from the average. Since the consumption system of a given country is reflected in the ecological footprint, according to our anticipation, the structure of the ecological footprint of countries, which are close to each other geographically and culturally, will also resemble each other. Mongolia and Uruguay form a stable and separated group because of their essential grazing land footprint, the year of study or the method notwithstanding. For this reason, we consider it reasonable to examine their consumption structure and characteristics in detail later. In case of the other three members group, the outstanding values can trace back to other reasons. Presumably, the consumption structures of Mauritius, African Gambia and North European Norway differ from each other considerably. The common feature means the fishing ground ecological footprint of 2010, which exceeds remarkably the average in all three countries. However, when examining the database of 2011, the fishing ground ecological footprint of Gambia is found corresponding to the world average. The reason for establishing cluster in 2010 can be attributed to the deviations in calculation methods of the fishing ground ecological footprint, which can strongly query the commensurability and reliability of the database. Our suggestion is that the statistical examination of the database (filtering the extreme outliers) should be followed by a professional control and the final data chart should be composed as a result of this. The consequences and the political decisions based on faulty and unreliable data can not lead to the expected results; it can make the situation even worse. The most essential result of our study can be a correction of failures in the database and we can establish a statistical background for more reliable consequences.

References

Bergh van den J.C.J.M., Verbruggen H. [1999]: Spatial sustainability, trade and indicators: an evaluation of the 'ecological footprint'. *Ecological Economics* no. 29 (1), pp. 61–72.
Ewing B., Reed A., Galli A., Kitzes J., Wackernagel M. [2010]: Calculation methodology for the National Footprint Accounts, 2010 Edition. Global Footprint Network, Oakland. [Available at:] http://www.footprintnetwork.org/images/uploads/National_Footprint_Accounts_Method_Paper_2010.pdf . [Accessed: May 2012].
Global Footprint Network. Downloadable database. [2010, 2011]. [Available at:] www.footprintnetwork.org. [Accessed: May 2012].
Kitzes J., Galli A., Rizk S., Reed A., Wackernagel M. [2008]: Guidebook to the National Footprint Accounts, 2008 Edition. Global Footprint Network. [Available at:] http://www.footprintnetwork.org/images/uploads/National_Footprint_Accounts_Method_Paper_2010.pdf. [Accessed: May 2012].
McDonald G.W., Patterson M.G. [2004]: Ecological footprints and interdependencies of New Zealand regions (analysis). *Ecological Economics* no. 50, pp. 49-67.
Sajtos L., Mitev A. [2007]: SPSS kutatási és adatelemzési kézikönyv[6]. Alinea Kiadó, Budapest.
Stiglitz J., Sen A., Fitoussi J.P. [2009]: Report by the Commission on the Measurement of Economic Performance and Social Progress. [Available at:] http://www.stiglitz-sen-fitoussi.fr/documents/rapport_anglais.pdf. [Accessed: May 2012].

[6] SPSS research and data analysis manual.