How many (and which) Indicators are necessary to compare the Environmental Performance of Companies? A sectoral and statistical answer

Paper accepted for The 7th European Roundtable on Cleaner Production, 2-4 May 2001, Lund, Sweden

Written by: Walter Wehrmeyer[1], Daniel Tyteca[2], Marcus Wagner[3]

The authors dedicate this paper to the memory of Jérôme Carlens, a former colleague who died under tragic circumstances in April 2001.

Abstract

Over the last 5 years, substantial progress has been made in the development and use of environmental performance indicators, so that, by and large, the problem of indicator availability has shifted to the more intricate problem of indicator suitability. Focusing on the need to compare companies' performances, the question today is "how many (and which) indicators are the minimum necessary to provide an approximate yet reasonably robust description of the comparative environmental performance of companies"? The results suggest that it is possible to provide a reasonably accurate picture of the total diversity contained in the data (i.e. diversity of firm performance) by using a minimum set of variables. This has obvious implications for management as well as public policy.

Keywords: Environmental performance indicators, electricity sector, paper sector, MEPI, Principal component analysis.

The first step is to measure whatever can easily be measured. This is OK as far as it goes. The second step is to disregard that which can't be easily measured or give it an arbitrary quantitative value. This is artificial and misleading. The third step is to presume that what can't be measured easily really isn't important. This is blindness. The fourth step is to say that what can't be easily measured really does not exist. This is suicide.

The MacNamara Fallacy (in Gray, 1993)

[1] Centre for Environmental Strategy; University of Surrey; GU2 5XH, Surrey, UK, Tel: +44 1483 259559, Fax: +44 1483 259394, E-mail: W.Wehrmeyer@surrey.ac.uk (Corresponding Author)

[2] Centre Entreprise - Environnement, Université Catholique de Louvain, Place des Doyens 1, B-1348 Louvain-la-Neuve, Belgium. Tel. +32.10.47.83.75, fax +32.10.47.83.24; E-mail tyteca@qant.ucl.ac.be.

[3] Lehrstuhl BWL, insb. Umweltmanagement, Scharnhorststr. 1, D-21335 Lüneburg, Germany, E-mail: mwagner@uni-lueneburg.de

Introduction

Over the last few years, a surprisingly large number approaches have sprung up to suggest, and subsequently apply, indicators that describe and inform about the environmental behaviour of organisations, sites, companies or manufacturing units. Earlier approaches typically focus on how such indicator systems should be designed and applied (e.g. (Wehrmeyer & Tyteca, 1998), (National Roundtable on the Environment and the Economy (NRTEE), 1997); (Ditz & Ranganathan, 1997); (Azzone & Noci, 1996) to name but a few). Lately, the debate has become more empirical and more detailed on the type of indicators that can be used for specific applications (e.g. (Commission of the European Communities (DGXI and XII) & Eurostat, 1996), (Gee & Moll, 1998); (Wright, Allen, Clift, & Sas, 1997)to name but a few). A significant move towards harmonising this diversity has been the publication and use of ISO14031 and ISO14032 in 1999, with more standards in preparation. Alongside this encouraging development, one could witness a sometimes bewildering patchwork of additional indicators of social or sustainability performance (GRI, 2000), of supplier, customer and end-user performance as well as of organisational performance that addresses special issues, such as biological diversity or community relations.

For this paper, focusing on environmental indicators for companies , two ideas can be drawn from the above. Firstly, which indicator is used does not any more depend on availability on cost criteria alone, but obviously depends on the context and purpose of the application, such as (a) reporting, (b) interpreting or (c) comparing environmental behaviour over time and between comparable entities, such as sites, companies or industry sectors. In addition, different users have different reasons to evaluate the environmental performance of a company and therefore may prefer different indicators. Secondly, the mushrooming of individual indicators can benefit greatly from standardisation and harmonisation. Such standardisation refers to the type of indicator itself as well as, perhaps more importantly, the ways in which data is to be gathered (including frequency, measurement units, measurement conditions and parameters) to allow in due course better data quality on a more comparable data set. As a result, the problem for decision-makers has shifted from "how can environmental performance of a company be measured" towards "which of these indicators do make most sense in given circumstances"?

An answer to this crucially depends on the level of detail and the degree of complexity allowed in the evaluation. For instance, if we were to compare many companies across different countries and even different sectors, the level of detail and comparability would inevitably be low and different indicators may have different meaning. In fact, some may argue that the level of detail is so low that it makes very little sense to do such an evaluation if detailed company recommendations are to be made. However, such an evaluation would probably yield important insights into the overall or general behaviour of firms in the environmental management field. For cost reasons, such an evaluation would also only include a limited set of indicators. We would have a - perhaps opaque - picture representing whole industries or countries.

By contrast, if we were to compare two very similar companies, the level of detail would be much deeper and we would be able to gain a much clearer and "sharper" picture of the environmental behaviour of the companies as not only would we be using more indicators, but also we could probably draw on indicators from many different years. However, the generalisation of the conclusions from this comparison would be very limited, as a sample of two is normally not representative at all. The question here can be specified as "how many (and which) indicators are the minimum necessary to give an approximate yet reasonably robust description of the comparative environmental performance of companies"? The underlying corollary to this is what environmental data is necessary to be collected for an appropriate comparison of companies. It should be noted that we do not advocate that the diversity of indicators be limited to a “general” set, but that there are circumstances in which not all indicators are needed (or their use is not feasible) and that in these circumstances, it would be helpful if the choice of indicators that are being used provide an approximate picture of the spread of environmental performance across the companies that are being studied. In other words, rather than restricting the development and diversity of indicators, we advocate that steps are in place to ensure that the indicators that are being used are representative for the overall diversity in environmental performance. Obviously, using a limited set of indicators will only provide a limited picture which will not allow a detailed examination of the root causes of the differences, nor of the more precise natures of that difference.

The paper reports on efforts to identify which indicators are essential in the description of environmental performance across a large set of companies. The data was collated as part of a European Commission funded project involving seven research partners across Europe covering six industrial sectors (Pulp & paper, Fertiliser, Textile finishing, Printing, Electricity Generation and Computer manufacturing) within six European countries (Austria, Belgium, Germany, Italy, Netherlands and the UK). Due to data availability this paper uses the sectors pulp and paper and Electricity generation as examples:

	Paper	Electricity
Firm-years	270	184
No of variables initially collected	88	58
Effective number of data points	10008	4482
Effective % missing values:	26%	29.3%

The structure of the paper is simple: after a brief section outlining the method being used, the paper presents the key findings for each of the sectors, followed by a discussion and conclusion, which outlines which indicators are suggested, as well as areas for future research.

The data / Method

Data was gathered as part of the MEPI project (www.environmental-performance.org) on the physical environmental performance of electricity generating companies and of paper producing companies. Data refers mainly to 1994-1998, although few cases were from 1985 and 1990. Distribution of sites and firms from both sectors across the participating countries is as follows:

	Number of firms			Number of production sites		
	Electricity	Paper	Sum	Electricity	Paper	Sum
Austria	9	8	17	13	8	21
Belgium	2	4	6	39	4	43
Germany	27	43	70	29	49	78
United Kingdom	10	8	18	52	18	70

Italy	6	10	16	20	12	32
Netherlands	4	17	21	9	16	25
All countries	58	90	148	162	107	269

Prior to data analysis, the data was first standardised per functional unit for each sector, followed by a review of the usefulness of some variables for the analysis. This covered the exclusion of some variables which simply had insufficient data available, and the review of outliers, which in some very limited cases were excluded. These stages are described in the section on Data Cleanup.

Description of Method

The approach to data analysis combines several value statements or observations with a numerical analysis of the empirical data. The value statements or observations that guided the analysis and the choice of statistical methods are:

1. The model implicit in the analysis perceives the organisation as a black box, with resources and energy as input, and emissions, waste energy and the product as output. As a result, the data sets for air and water emissions, as well as for waste production have been analysed separately. From a mass balance perspective, it does not matter whether inputs or outputs are analysed, from the perspective of access to data, outputs have been focused.

2. Many environmental variables have substantial correlations between them (multicollinearity). For example, there is a strong correlation between CO2, SO2 and NOX in fossil-fuel fired power stations (r^2 is between 0.917 and 0.982; n p=001, n=118 and 116), or we can find a strong correlation between BOD and COD in water emissions of paper companies (r^2 = 0.878, p=0.01; n=118). As a result, redundant variables can be excluded for the benefit of one variable that represents a set of highly correlated other variables.

3. Inasmuch as it makes little sense to compare companies' physical environmental performance across very different sectors (such as paper and pulp production and electricity), it does not make sense to integrate inputs and outputs across all media (air, water and soil) into one large indicator. It was decided not to seek this one large indicator for the benefit of a better analysis of the comparison between companies in the same sector. The same argument can hold for single-indicator ranking systems. This represents a preference of the researchers.

The method that was primarily used to identify statistically robust results is that of principal component analysis, which is designed to reduce the number of variables in a given data set to simplify with minimal loss of information (Everett; 1974; p. 4). Principal Component Analysis intends that variances of most of the (principal components) will be so low as to be negligible. In that case the variation in the data set can be adequately described by the few (principal components) with variances that are not negligible (Manly; 1986, p. 58). Thus, Principal Component Analysis can be seen as a useful linear tool to reduce the large number of variables to a manageable number by transforming the variables into components which are independent from each other, so-called eigenvectors (Anton, 1991; Everett, 1974). Factor analysis is often used in data reduction to identify a small number of factors that explain most of the variance observed in a much larger number of manifest variables. Factor analysis can also be used to generate hypotheses regarding causal mechanisms or to screen variables for subsequent analysis (for example, to identify collinearity prior to performing a linear regression analysis).

Regressions try to predict one variable by a combination of one or several others. PCA tries to identify new variables that retain the diversity of responses, but can express this with a minimised (or optimised) number of variables. The main reasons for using this method are:

- Its underlying assumptions are fairly robust. For instance, different to the most common other method, regression, it does not rely on an underlying model that assumes whether the relationship between variables is linear, logarithmic, quadratic etc, It is an optimisation method that is based on matrix calculations based on the correlation coefficients of each variable with all others.
- It is easy and simple to use;
- It can reduce large numbers of variables in a way, that allows an assessment of the quality of the variable reduction.
- The treatment of missing values can be controlled very well, which is of particular concern here, given that there are so many missing values in the dataset.

Data Cleanup

The initial data was standardised per functional unit, which was in the electricity sector kW hours produced (an output measure) and in the paper industry metric tonnes of paper produced. Reducing the general problem outlined above to an operational problem, the question this paper addresses is, given a set of environmental data with a large number of variables and many missing values, which indicators are particularly useful in describing the diversity of firm behaviour? The initial dataset of normalised variables covered these indicators:

Number of Cases for each Variable in both Sectors

	Paper	Electr.		Paper	Electr.
Additives per FU	46		Nitrogen per FU	84	5
AoX per FU	31		Ink Input per FUK	6	
BOD per FU	82	2	PM10 per FU	3	7
CO2 per FU	63	118	publ H2O supply per FU	72	6
Coal Input per FU	65	71	Rec Fibre per FU	102	
COD per FU	107	8	Recycl Waste per FU	71	56
Dust emissions per FU		60	Renewables inp per FU	68	20
Electr Tot inp per FU	54		Phosph per FU	54	4
Ext Electr per FU	101	14	VOC per FU	18	8
Gas input per FU	144	86	SO2 per FU	44	135
H2O Extract per FU	106	5	Tot Electr inp per FU	64	16
Haz waste per FU	71	58	Tot energy per FU	39	10
Heavy Metals per FU	55	2	Tot Fuel Input per FU	73	16
Mun Waste per FU	95	59	Tot H2O per FU	120	9
NOx per FU	117	134	Tot sol Waste per FU	53	74
Spent fuel per FU		9	Transp Fuel per FU	6	
On-site Electr inp per FU	35	4	Tot oil inp per FU	73	78
Available Firm-years	270	184			

No firm in either sector provided a complete dataset. In fact, as not all companies provided data about the functional unit, the standardisation process alone reduce the available data set. As can be seen from the above table, a lot of data could be gathered, but there are also substantial gaps in the dataset. Accordingly, the treatment of missing values is a critical factor in the data analysis. In addition, a cross-sectoral analysis is virtually impossible, as, with few notable exceptions, indicators that are data-rich in one sector are very sparse in the other sector. This makes intrinsic sense as the environmental performance of each sector is fundamentally different, so that such a comparison would “compare apples and pears”. As a result, such a comparison is not being tried, but both sectors are being used as stand-alone sector industries that are used to evaluate the suitability of the method.

Finally, there are some variables that are so underrepresented that their use is, for this analysis, unsafe. For instance, particulates per functional unit (PM10 per FU) has only data for 3 out of 270 firm-years (essentially one company reporting particulates data over three years). There is not enough data available to even begin including such variables as, in this case, 98.8% of data is missing.

Results from the Paper Sector

In this section, the results for the paper industry are outlined. To explain the method, the results for waste are explained in more detail.

Waste

The initial PCA for all Waste variables (below) shows that there are two components that in total explain 79.8% of the total variability contained in the waste variables can be explained using the two factors (cumulative proportion). The first factor can explain 52.2%

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.089	52.237	52.237	2.089	52.237	52.237
2	1.103	27.587	79.824	1.103	27.587	79.824
3	.800	19.997	99.821			
4	7.168E-03	.179	100.000			

Extraction Method: Principal Component Analysis.

In it, the first factor is essentially a combination of total solid waste and the amount of recycled waste; the second factor is a combination of total hazardous waste and municipal waste. This result was independent of the way missing values were treated.

Component Matrix[a]

	Component	
	1	2
Tot sol Waste per FU	.976	-7.71E-02
Haz waste per FU	-.182	.808
Mun Waste per FU	.360	.666
Recycl Waste per FU	.987	-1.78E-02

Extraction Method: Principal Component Analysis.

a. 2 components extracted.

Together, the two tables suggest that just over half of the variability can be explained by only using total solid waste and recycled waste. Hence, using the variables “Municipal Waste” or “Hazardous Waste” (per functional unit) may add further insights into company behaviour, but were found not to be those variables that are the most important ones in the description of firm behaviour.

Air and water emissions

Availability of data on air emissions for the paper sector was relatively, poor and most variables had to be excluded due to seriously missing variables. The remaining three variables for which sufficient data was available showed inconclusive results in the PCA. Depending on the way missing values are treated, the proportion of data variability that can be explained varies, and so do the factor loadings – the relative weight of the variables in explaining the data. Below are the results for three individual PCAs highlighting the problem:

	PCA1	PCA2	PCA3
Missing value treatment	Listwise deletion[1]	Pairwise deletion[2]	Replacement with mean[3]
Proportion explained by the factor	63.31%	56.06%	40.64
CO2 per FU	0.787	0.871	0.781
NOX per FU	0.751	0.571	0.781
SO2 per FU	0.847	0.772	0.00254

The above table also shows that the importance of the first Factor typically falls with a more relaxed way of treating missing values. This is because the number of cases that are included rises. However, given the high correlations between CO2 and SO2, it seems reasonable that CO2 is an adequate representation for air emissions if a rough estimate is desired.

With regard to water emissions, data quality was much better. It was found that the majority of the data diversity (70%) can be represented using COD, Nitrogen and Phosphorous per FU.

Initially, 32 variables were collected from the paper industry. Of these, a number were so poorly represented that they could not be used for a meaningful and statistically robust data analysis. However, when grouped into emissions and wastes, 6 appeared to be sufficient to describe the waste, air and water emissions adequately for an inter-firm comparison. These variables are total solid waste, recycled waste, CO2, COD, Nitrogen and Phosphor per FU. If total energy consumption and total water consumption is added to this list, it appears that a reasonable picture of firm behaviour can be identified using 8 rather than 32 variables. This represents a significant cost advantage in the gathering, storing and interpreting environmental data.

Results from the Electricity Sector

Using the same approach as for the paper industry, the respective Factors (variables) and their proportion of the variance explained are as follows:

Environmental Media	Variables	Approximate proportion of variance explained by Factor
Waste	Total Waste and Recycled Waste	45-70%
Air Emissions	CO2, SO2, NOx,	60%
Water emissions	Data availability was too poor to warrant an analysis	
Energy	Total Fuel input	57%
Water Consumption	Data availability was too poor to warrant an analysis	

Again, it was found that of the majority of variables, only few are necessary to gain an approximate picture of the environmental performance of the company.

[1] One missing value in either variable means the firm-year is excluded from the analysis.
[2] Excludes cases with missing values for either or both of the pair of variables in computing a particular statistic.
[3] Replaces the missing value with the mean for the variable.

Discussion and Conclusions

It appears that, to gain an approximate rather than a precise picture, much fewer variables are needed to describe the environmental performance of companies in these two sectors. This is at first glance a very encouraging conclusion, as efforts to gain more variables may under certain circumstances be replaced (to save costs) or redirected towards more comprehensive collections of companies.

However, the main criticism against this is based on the observation that the variables that are recommended here as a set of variables that suffice for broad comparisons between companies in the same sector are also those that are the most common ones. From a methodological viewpoint, there may be variables which are even better in describing or comparing corporate environmental performance, but were not included either because these variables are not known as yet, or because insufficient data was collated for their inclusion. This may well be, and, regarding the first possibility, until technical knowledge is available that supports such variable(s), we would have to suffice with the data set as given.

With regard to the second one, the assumption that richly available variables are an indication that these data are useful within the industry is a reasonable one. Both sectors have several decades of environmental exposure, the debate about the environmental efficacy of fossil fuels is at least 35 years old, and may well go back to the industrial revolution. It is reasonable to suggest that engineers and policy makers have had enough time to identify the most significant environmental variables for inclusion and, as a result, are also those that are collated by the industry more frequently than others.

But there is the danger that this empirical research that recommends variables based on their explanatory power does recommend these based on their availability. This is possible, although a number of variables are not suggested here even though data is available (e.g. SO2 or NOx) based on the observation that these variables can be represented by others. However, this is also unlikely as for most PCAs, the availability of data (as expressed by the use and treatment of missing values) has not changed the results significantly.

In conclusion, the empirical analysis suggests that if we are satisfied with an approximate rather than a precise analysis, one may use far less variables than otherwise suggested. In fact, those variables that are suggested here are also fairly common throughout either industry. This is not to suggest that for more specific questions, other variables are much more important, but in a large-firm comparison, the representation of the diversity of firm behaviour across 32 variables is not that different than across 8 variables. This suggest great savings in the cost of data collection as we can be reasonably robust in the comparison of firms across fewer variables. The result of this research is also to some degree putting into question current activities (such as the Global Reporting Initiative) which attempt to specify an elaborate set of environmental, social and economic performance indicators to be used compulsory by firms when assessing sustainability. The major contribution of such initiatives (apart from establishing what should be measured) may well be a consensus regarding how indicators should be measured, i.e. precise and consistent data collection protocols for each indicator. Given that the larger the number of indicators, the more difficult it will be a) to reach consensus if and how to include all indicators, and b) to establish detailed protocols for each of them, activities like the Global Reporting Initiative (GRI) may have a long way ahead. The longer the time to consensus, the higher the threat that initiatives and activities like GRI loose momentum before significant achievements have been made. From this perspective research and analyses like the one presented can assist a pre-selection of indicators and thus provide valuable input into activities aimed at standardising indicator sets for environmental (as well as social) performance measurement.

References

Azzone, G., & Noci, G. (1996) Defining Environmental Performance Indicators: An Integrated Framework. *Business Strategy and the Environment*, Vol. 5, 69-80.

Commission of the European Communities (DGXI and XII), & Eurostat. (1996) *Environmental Indicators and Green Accounting*. Brussels: Commission of the European Communities.

Ditz, D., & Ranganathan, J. (1997) *Measuring Up: Toward a Common Framework for Tracking Corporate Environmental Performance*. Washington: World Resources Institute.

Gee, D., & Moll, D. (1998) *Information for Sustainability: Eco-Efficiency Indicators*. Kopenhagen: European Environment Agency (EEA).

Global Reporting Initiative (2000), http://www.globalreporting.org/Guidelines/June2000/SupportingDocuments /FacilityLinking.htm, accessed 19 January 2001

National Roundtable on the Environment and the Economy (NRTEE). (1997) *Backgrounder: Measuring Eco-efficiency in Business*. Ottawa, Canada: National Roundtable on the Environment and the Economy .

Wehrmeyer, W., & Tyteca, D. (1998) Measuring Environmental Performance for Industry: From Legitimacy to Sustainability? *The International Journal of Sustainable Development and World Ecology*, Vol. 5, 111-124.

Wright, M., Allen, D., Clift, R., & Sas, H. (1997) Measuring Corporate Environmental Performance: The ICI Environmental Burden System. *Journal of Industrial Ecology*, Vol. 1(4), 117-127.