

Northumbria Research Link

Citation: Ozoemena, Matthew, Cheung, Wai Ming, Hasan, Reaz and Hackney, Philip (2014) A hybrid Data Quality Indicator and statistical method for improving uncertainty analysis in LCA of a small off-grid wind turbine. In: ARCOM Doctoral Workshop on Sustainable Urban Retrofit and Technologies, 19 June 2014, London South Bank University.

URL:

This version was downloaded from Northumbria Research Link: <http://nrl.northumbria.ac.uk/16687/>

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: <http://nrl.northumbria.ac.uk/policies.html>

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher's website (a subscription may be required.)



**Northumbria
University**
NEWCASTLE



UniversityLibrary

A hybrid Data Quality Indicator and statistical method for improving uncertainty analysis in LCA of a small off-grid wind turbine

Matthew Ozoemena, W.M. Cheung, Reaz Hasan and Phil Hackney

Department of Mechanical and Construction Engineering, Faculty of Engineering and Environment, Northumbria University, Newcastle Upon Tyne, NE1 8ST, UK.

In Life Cycle Assessment (LCA) uncertainty analysis has been recommended when choosing sustainable products. Both Data Quality Indicator and statistical methods are used to estimate data uncertainties in LCA. Neither of these alone is however adequate enough to address the challenges in LCA of a complex system due to data scarcity and large quantity of material types. This paper applies a hybrid stochastic method, combining the statistical and Data Quality Indicator methods by using a pre-screening process based on Monte Carlo rank-order correlation sensitivity analysis, to improve the uncertainty estimate in wind turbine LCA with data limitations. In the presented case study which performed the stochastic estimation of CO₂ emissions, similar results from the hybrid method were observed compared to the pure Data Quality Indicator method. Summarily, the presented hybrid method can be used as a possible alternative for evaluating deterministic LCA results like CO₂ emissions, when results that are more reliable are desired with limited availability of data.

Keywords: CO₂ emission, data quality indicator, lca, statistical, monte carlo simulation.

INTRODUCTION

Estimating CO₂ emissions is a significant part of wind energy LCA's. Traditionally CO₂ emission is estimated with a deterministic approach which uses a fixed point value to represent emission factor and generate a single fixed point result. Due to differences in emission factors which may vary by industrial process (Wang and Sun, 2012), there could be significant variations in emission factors among different life cycle inventory (LCI) databases. These variations can affect the results of CO₂ emissions significantly. Incorporating the analysis of data uncertainty of emission factors could be an important improvement to the deterministic approach as it can provide more information for decision making. According to Wang and Sun (2012), CO₂ emission is given by the following formula:

$$\begin{aligned} \text{Emission} &= \sum_{i=1}^n \text{Emission}_i \\ &= \sum_{i=1}^n \text{Activity Level}_i \times \text{Emission Factor}_i \end{aligned} \quad (1)$$

Where,

Emission i: Amount of CO₂ emitted from the consumption of material i (e.g. iron)

Activity level i: Material consumption for material i

Emission factor i: Consumption of material i's emission factor

Data quality indicator (DQI) and statistical methods are often used to estimate data uncertainty in LCA with differing shortcomings and advantages (Sugiyama et al., 2005; Junnila and Horvath, 2003; EPA, 1995; Hanssen and Asbjørnsen, 1996). DQI estimates the uncertainty and reliability of data based on expert knowledge and descriptive metadata such as the data's completeness, geographical correlation, etc. It is mentioned in Coulon et al., (2011) and Junnila and Horvath (2003) that DQI can be used both quantitatively and qualitatively in LCA studies. On the other hand, the statistical method fits data samples with a goodness of fit test to characterize data range with probabilistic distributions if enough data samples are available. DQI although less accurate than the statistical method costs less compared to the statistical method (Venkatesh et al., 2011; Tan et al., 2002b). Due to the high cost of implementing the statistical method, though it is desirable when high accuracy is required, DQI is extensively applied when high accuracy of an uncertainty estimate is not critical or the size of a data sample is not large enough for meaningful statistical analysis (Sugiyama et al., 2005). Considering the trade-offs between cost of implementation and accuracy, Wang and Shen (2013) presented an alternative stochastic solution using a hybrid DQI-statistical (HDS) approach to improve the quality of pure DQI method while reducing the cost of the pure statistical method in whole-building LCA. The key departure from previous works being the stochastic pre-screening process using quantitative DQI and Monte Carlo simulation (MCS) to determine the influence of the contribution of parameters. After the categorization, the statistical method is adopted for the critical parameters, and DQI based distributions are used for non-critical parameters. An application test case to wind turbine LCA is presented to validate the presented solution. The aim of this paper is to present the hybrid DQI-statistical (HDS) method to improve the uncertainty estimate of CO₂ emissions of a small off-grid wind turbine combining the advantages of the traditional DQI and the statistical method to develop a more practical approach. This method can be used as a valuable tool to evaluate deterministic results of CO₂ emissions when uncertainty information is needed for decision making.

METHODOLOGY

The DQI Method

DQI characterizes the quality of data using descriptive indicators often formatted as a data quality pedigree matrix (DQPM) as shown in Table 1. Columns in the matrix represent data quality indicators such as data's completeness, age etc. while rows represent the quality scale from 1 – 5. The overall quality of data can be characterized by an aggregated number that takes into account all the individual indicators (Junnila and Horvath, 2003). All the indicators are treated equal in weight, for example, if (5, 4, 3) are assigned to three indicators respectively, the aggregated DQI score for the parameter is $T = 5 \times 1/3 + 4 \times 1/3 + 3 \times 1/3 = 1.61$.

Quantitative DQI

Quantitative DQI enables the transformation of aggregated DQI scores to probability density functions for the quantification of uncertainty (Weidema and Wesnæs, 1996; Tan et al., 2002b, Maurice et al., 2000; May and Brennan, 2003). The idea being to characterize data of different quality by probability density functions based on the “rule of thumb” (Finnveden and Lindfors, 1998). The DQI transformation matrix is often used to convert aggregated DQI scores into beta functions (May and Brennan,

2003; Canter et al., 2002; Tan et al., 2002b; Kennedy et al., 1997; Kennedy et al., 1996).

$$f(x; \alpha, \beta, a, b) = \left[\frac{1}{b-a} \right] * \left\{ \frac{\Gamma(\alpha + \beta)}{[\Gamma(\alpha) * \Gamma(\beta)]} \right\} * \left[\frac{x-a}{b-a} \right]^{\alpha-1} * \left[\frac{b-x}{b-a} \right]^{\beta-1} \quad (2)$$

$$(a \leq x \leq b)$$

Where α, β are distribution shape parameters and a, b are selected range endpoints.

Canter et al. (2002) suggests the use of the beta function due to the fact that “shape parameters and range end points allow virtually any shape probability distributions to be represented”. As expressed by Canter et al., (2002), “the shape parameters establish the shape of the distribution and thus the location of the probability mass, whereas the endpoints limit the range of possible values”.

HDS Approach

Wang and Shen (2013) states that the HDS approach consists of four steps: (a) Quantitative DQI with MCS; (b) Parameter characterization; (c) Detailed probability distributions estimation for parameters; and (d) Final MCS calculation. The parameter characterization identifies the critical parameters based on the parameters’ degree of uncertainty and their influences. The final stochastic results will be produced through a MCS calculation.

Table 1: Data quality pedigree matrix (DQPM) based on National Energy Technology Laboratory (NETL) LCI&C Guideline Document

Indicator	Score				
	1	2	3	4	5
Source Reliability (for most applications, source quality guidelines only factor)	data verified based on measurements	data verified based on some assumptions and/or standard science and engineering calculations	data verified with many assumptions, or non-verified but from quality source	qualified estimate	non-qualified estimate
	source quality guidelines met		source quality guidelines not met		
	data cross checks, greater than or equal to 3 quality sources	2 or less data sources available for cross check, or data sources available that do not meet quality standards		no data available for cross check	
Completeness	representative data from a sufficient sample of sites over an adequate period of time	smaller number of site but an adequate period of time	sufficient number of sites but a less adequate period of time	smaller number of sites and shorter periods or incomplete data from an adequate number of sites or periods	representativeness unknown or incomplete data sets
Temporal Correlation	less than three years of difference to year of study/current year	less than 6 years of difference	less than 10 years difference	less than 15 years difference	age of data unknown or more than 15 years difference
Geographical Correlation	data from area under study	average data from larger area or specific data from close area	data from area with similar production conditions	data from area with slightly similar production conditions	data from unknown area or area with very different production conditions
Technological Correlation	data from technology, process or materials being studied	data from a different technology using the same process and/or materials		data on related process or material using the same technology	data or related process or material using a different technology

a) Quantitative DQI with MCS

This step follows Canter et al. (2002)'s methodology beginning with data quality assessment using DQI. All parameters used for the deterministic calculations are evaluated based on the DQPM. After the calculation of aggregated DQI scores, probability distributions for each of the parameters are estimated based on the transformation matrix (Table 2), and used as inputs for the MCS to perform an influence analysis.

b) Parameter characterization

The degree of parameter uncertainty can be obtained in the process of data quality assessment. Accordingly, parameters will be classified into groups of four with DQI scores belonging to the intervals of (Alcorn and Baird, 1996; Ortiz et al., 2009), (4, 5), (3, 4), (2, 3) and (1, 2) respectively. The group containing parameters with DQI scores within the interval of (1, 2) and (2, 3) show the highest uncertainty, and the group with parameters scored within the (3, 4) and (4, 5) interval depict the highest certainty. Sugiyama et al. (2005) shows that a parameter's influence on the final resulting uncertainty comes from a rank-order correlation analysis in MCS (Equations (3) and (4)).

$$IA_{p,q} = r_{p,q}^2 \left[\sum_p r_{p,q}^2 \right]^{-1} \times 100\% \tag{3}$$

Where $IA_{p,q}$ is the influence of input parameter p to output q ; $r_{p,q}$ is the rank-order correlation factor between input p and the output q . $r_{p,q}$ can be computed via:

$$r_{p,q} = 1 - \left[6 / (N^3 - N) \right] \sum_{i=1}^N [rank(p_i) - rank(q_i)]^2 \tag{4}$$

Where $rank(p_i)$ and $rank(q_i)$ are the ranks of p_i and q_i among the N tuple data points.

Table 2: Transformation matrix based on (Canter et al., 2002 and Weidema and Wesnæs, 1996).

Aggregated DQI scores	Beta distribution function	
	Shape parameters (α, β)	Range endpoints (+/- %)
5.0	(5, 5)	10
4.5	(4, 4)	15
4.0	(3, 3)	20
3.5	(2, 2)	25
3.0	(1, 1)	30
2.5	(1, 1)	35
2.0	(1, 1)	40
1.5	(1, 1)	45
1.0	(1, 1)	50

c) Detailed probability distributions estimation for parameters

The statistical method will be applied, after the parameter categorization, to the process of fitting probability distributions of the identified critical parameters. Kolmogorov-Smirnov goodness of fit test (K-S test) is a statistical tool that can be

applied for determining whether a data sample is drawn from a population with a specifically hypothesized distribution by measuring the maximum vertical distance between the two cumulative distribution functions (Massey, 1951). If this distance is smaller than the designated critical table value, the null hypothesis that “The data sample follows the hypothesized distribution” can be accepted (Massey, 1951). The K-S test statistic is defined as:

$$D = \max_{1 \leq i \leq N} \left[F(Y_i) - \frac{i-1}{N}, \frac{i}{N} - F(Y_i) \right] \quad (5)$$

Where F is the theoretical cumulative distribution of the distribution being tested; N meaning N ordered data points Y1, Y2 ... Yi ..., YN. For the non-critical parameters of lower uncertainty and influence, the probability distribution will be estimated based on the DQI scores and the transformation matrix.

d) Final MCS calculation

The final step is calculating the stochastic results by MCS algorithm, according to the relationship between inputs and outputs, using the elaborately estimated parameter probability distributions as inputs. The probability distributions of non-critical parameters are obtained from the quantitative DQI.

Validation

To validate benefits of the HDS, it is compared to the pure DQI using two measurements to measure the difference between the results. Mean Magnitude of Relative Error (MRE) (Eq. (6)) (Abdou et al., 2004) and Coefficient of Variation (CV) (Eq. (7)) (Venkatesh et al. 2010). A large CV value indicates wide spread of a distribution.

$$MRE = \frac{(M_{HDS} - M_{DQI})}{M_{HDS}} \times 100\% \quad (6)$$

Where MHDS is the mean of the HDS results and MDQI is the mean of the pure DQI results

$$CV = SD/M \quad (7)$$

Where SD is the standard deviation and M is the mean.

TEST CASE RESULTS AND DISCUSSION

Estimation of the CO2 emissions for three unit processes (Produce Air-X-9, Produce Tower and Produce Batteries), out of six, of a wind turbine LCA test case adopted from Fleck and Huot (2009) was performed. The reason only three of the processes were considered is in a large part, due to time constraints regarding the deadline for the submission of this paper. Since the quantities of the wind turbine components were from the same data source, they have very little or no variations. The deterministic estimate result was used as a benchmark for comparison of the stochastic estimation outputs.

Quantitative DQI transformation

Aggregated DQI scores were rounded off to the nearest nominal value in order to use the transformation matrix. Figure 1 shows the aggregated DQI scores. Because most of the parameters used in this test case were adopted from the same data source they showed the same DQI score of 4 and the same transformation beta function

parameters ($\alpha = 3, \beta = 3$), with the exception of battery and galvanized steel with DQI scores of 3.5 and 3 respectively.

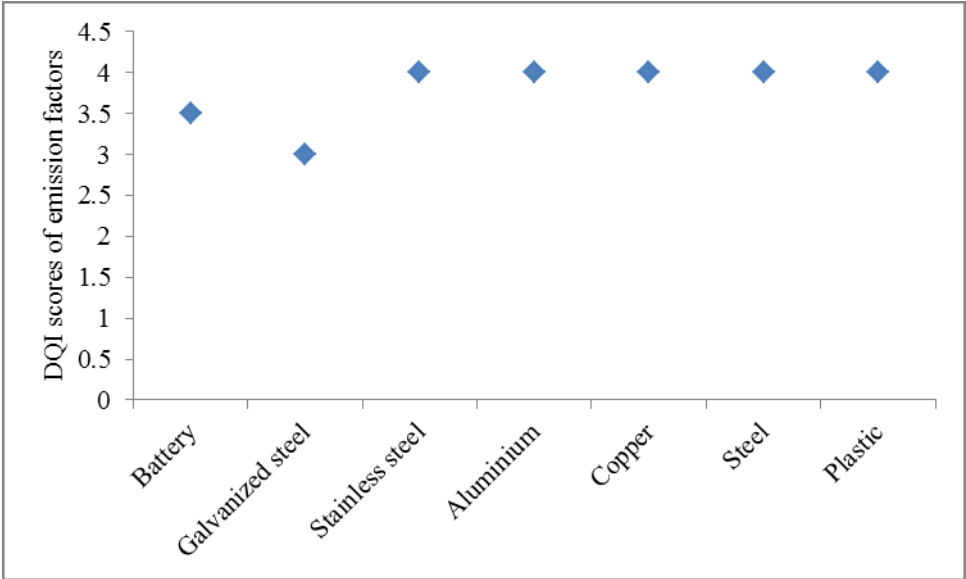


Figure 1: Aggregated DQI scores

Categorizing Parameters

The influence analysis results (2,000 runs MCS) are shown in figure 2. Aluminium emission factor shows the largest influence contributing 25% of the resulting uncertainty. The following parameter is plastic emission factor, contributing 21% of the resulting uncertainty. Majority of the data are of good quality with corresponding DQI scores of 4. The parameter galvanized steel emission factor is the most uncertain with a DQI score of 3. With these results aluminium emission factor and plastic emission factor were positioned for further analysis using the statistical method, while others obtained their values from the quantitative DQI.

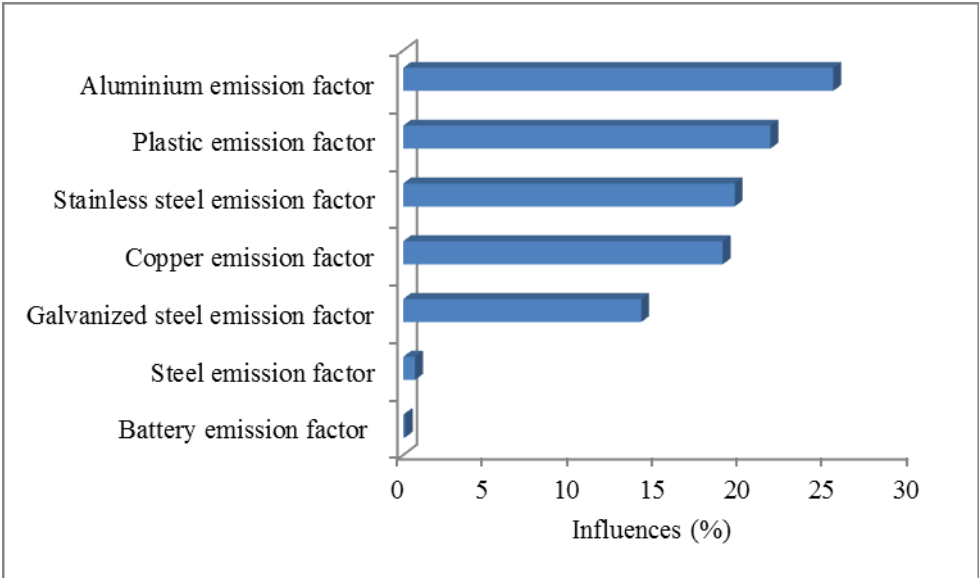


Figure 2: Influence Analysis

Probability Distributions Estimation

Beta (4.5, 5.2) was fitted to 32 data points manually collected from previous studies for aluminium emission factor with a mean value of 11.58 kg CO₂eq/kg. While for plastic emission factor, beta (1.8, 11.3) was fitted to 33 data points manually collected from literature with a mean value of 3.8 kg CO₂eq/kg.

Comparison of the HDS, Pure DQI and Deterministic Results

Figure 3 shows the stochastic result (2,000 runs MCS) using DQI. Beta distribution (4.5, 4) (K-S test) was fitted, with a mean value of 3531 kg CO₂eq and standard deviation of 401 kg CO₂eq. The HDS result follows a beta distribution (6.9, 9.7) (K-S test), with a mean value of 3535 kg CO₂eq and standard deviation of 327 kg CO₂eq. Thus, there is little difference in the dispersion from the DQI result. The CV value of the HDS result is 0.09, about 81% less than the value of 0.11 for the pure DQI result. The (10%, 90%) certainty interval for the output of the DQI is (3,032 kg CO₂eq, 4,083 kg CO₂eq) with a span of 1051 kg CO₂eq, while a slightly narrower (10%, 90%) certainty interval of (3,117 kg CO₂eq, 3,961 kg CO₂eq) with a span of 844 kg CO₂eq is presented for the HDS result. In terms of MRE, 0.11% difference was observed between the HDS and pure DQI result. This indicates that HDS, given the scope of this study, does not capture more possible outcomes than pure DQI, i.e. pure DQI does not underestimate the uncertainty of the result. The differences between the three results (deterministic, pure DQI and HDS) can also be seen from the cumulative distribution function. As seen in Figure 4, it can be concluded that about 50% of the possible results are smaller than the obtained deterministic result based on the HDS and pure DQI result curves. From the procedure of HDS which identifies critical parameters and handles them with the statistical method, which is presumed accurate, it can be seen that the final results generated from HDS are somewhat jeopardized. Since the identified critical parameters that explained the majority of the overall uncertainty was around 46%, it can be hypothesized that there is not much uncertainty in the data related to these processes given the little differences in the influences between the parameters. Consideration of the three remaining transport processes, where the data might have significant scatter, could meaningfully influence the result.

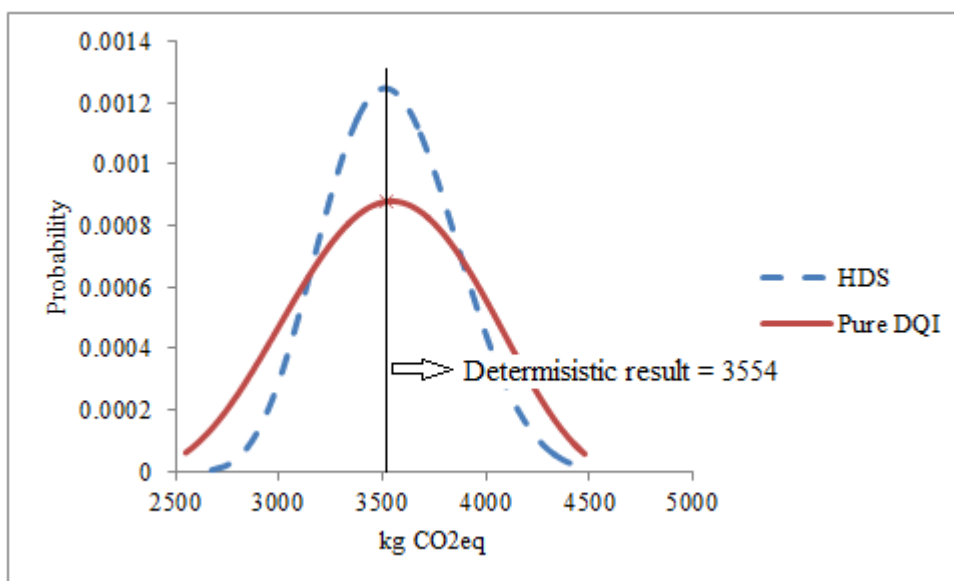


Figure 3: Comparison of resulting probability distributions between HDS and pure DQI

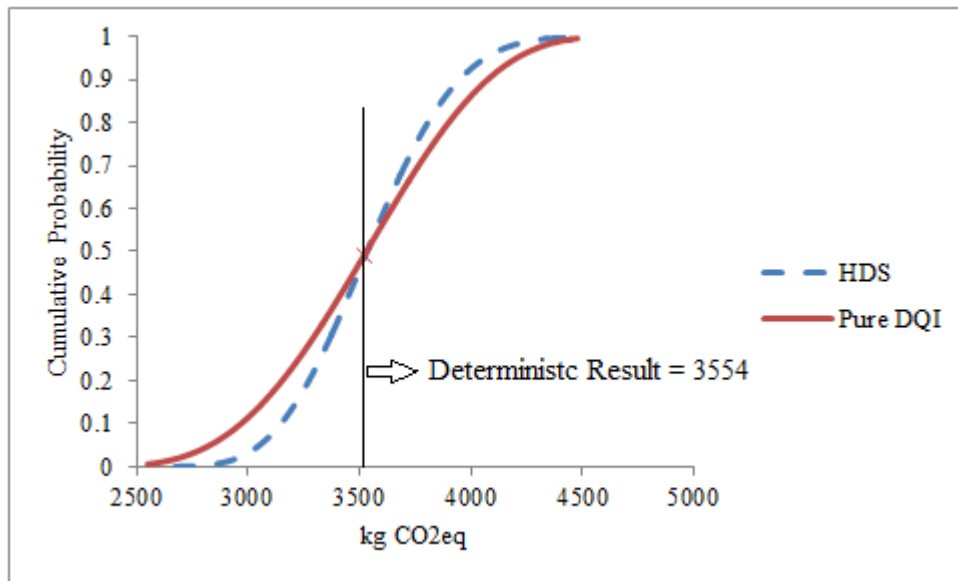


Figure 4: Comparison of cumulative probability results between HDS and pure DQI

CONCLUSIONS

The presented hybrid approach using a pre-screening technique based on Monte Carlo rank-order correlation sensitivity analysis did not demonstrate its effectiveness in evaluating deterministic results of CO₂ emissions emitted. The quantitative DQI method did not underestimate the data uncertainties compared to the HDS, which used the statistical method to estimate the most influential parameters. The results measured by MRE and CV between both methods indicate that HDS did not capture a wider range of uncertainties when compared to pure DQI. Evaluating the reliability of the deterministic value of CO₂ emissions, HDS did not show improved estimate of data uncertainties compared to DQI, meaning HDS approach did not mitigate the uncertainty underestimation deficiency of DQI. From Figure 4 it can be seen there is about 50% chance that the deterministic result is greater than the actual value using both methods. Thus decisions based on either approach are reliable.

REFERENCES

- Abdou, A., et al. (2004) Modelling risk for construction cost estimating and forecasting: a review. "Proceedings of the 20th Annual ARCOM Conference", Heriot Watt University, Edinburgh, September, ARCOM, Reading.
- Alcorn, J. and G. Baird (1996) Use of a hybrid energy analysis method for evaluating the embodied energy of building materials. "Renewable Energy", 8(1), 319-322.
- Canter, K. G., et al. (2002) Screening stochastic life cycle assessment inventory models. "The International Journal of Life Cycle Assessment", 7(1), 18-26.
- Coulon, R., et al. (1997) Data quality and uncertainty in LCI. "The International Journal of Life Cycle Assessment", 2(3), 178-182.
- EPA, April 1995 Guidelines for Assessing the Quality of Life-cycle Inventory Analysis. United States Environmental Protection Agency. EPA530-R-95-010.
- Finnveden, G. and L.-G. Lindfors (1998) Data quality of life cycle inventory data—rules of thumb. "The International Journal of Life Cycle Assessment", 3(2), 65-66.

- Fleck, B. and M. Huot (2009) Comparative life-cycle assessment of a small wind turbine for residential off-grid use. "Renewable energy", 34(12), 2688-2696.
- Hanssen, O. J. and O. A. Asbjørnsen (1996) Statistical properties of emission data in life cycle assessments. "Journal of cleaner production", 4(3), 149-157.
- Junnila, S. and A. Horvath (2003) Life-cycle environmental effects of an office building. "Journal of Infrastructure Systems", 9(4), 157-166.
- Kennedy, D.J., Montgomery, D.C., Quay, B.H., (1996) Stochastic environmental Life Cycle Assessment modeling: a probabilistic approach to incorporating variable input data quality. "Int. J. LCA", 1 (4), 199-207.
- Kennedy, D.J., Montgomery, D.C., Rollier, D.A., Keats, B.J., (1997) Data quality: assessing input data uncertainty in life cycle assessment inventory models. "Int. J. LCA", 2 (4), 229-239.
- Massey, F.J. (1951) "The Kolmogorov-Smirnov test for goodness of fit". J. Am. Stat. Assoc. 46 (253), 68-78.
- Maurice, B., et al. (2000) Uncertainty analysis in life cycle inventory. Application to the production of electricity with French coal power plants. "Journal of cleaner production", 8(2): 95-108.
- May, J. R. and D. J. Brennan (2003) Application of data quality assessment methods to an LCA of electricity generation. "The International Journal of Life Cycle Assessment", 8(4), 215-225.
- NETL (2010) NETL Life Cycle Inventory Data – Unit Process: Horizontal Turbine Main Frame, 1.5-6 MW Capacity, Manufacturing. U.S. Department of Energy, National Energy Technology Laboratory. Last Updated: November 2010 (version 01). www.netl.doe.gov/energy-analyses (<http://www.netl.doe.gov/energy-analyses>)
- Ortiz, O., et al. (2009) Sustainability in the construction industry: A review of recent developments based on LCA. "Construction and Building Materials", 23(1): 28-39.
- Sugiyama, H., et al. (2005) Using Standard Statistics to Consider Uncertainty in Industry-Based Life Cycle Inventory Databases. "The International Journal of Life Cycle Assessment", 10(6), 399-405.
- Tan, R. R., et al. (2002) Application of possibility theory in the life-cycle inventory assessment of biofuels. "International Journal of Energy Research", 26(8), 737-745.
- Venkatesh, A., et al. (2010) Uncertainty analysis of life cycle greenhouse gas emissions from petroleum-based fuels and impacts on low carbon fuel policies. "Environmental science & technology", 45(1), 125-131.
- Wang, Y. and T. Sun (2012) Life cycle assessment of CO2 emissions from wind power plants: Methodology and case studies. "Renewable Energy", 43, 30-36.
- Wang, E. and Z. Shen (2013) A hybrid Data Quality Indicator and statistical method for improving uncertainty analysis in LCA of complex system: application to the whole-building embodied energy analysis. "Journal of cleaner production", 43, 166-173.
- Weidema, B. P. and M. S. Wesnæs (1996) Data quality management for life cycle inventories—an example of using data quality indicators. "Journal of cleaner production", 4(3), 167-174.