Modeling the inventory of hydropower plants

Vincent Moreau[1,*], Gontran Bage[2], Denis Marcotte[3] and Réjean Samson[1]

[1]Ecole Polytechnique de Montreal, Department of Chemical Engineering, International Chair in LCA, Montreal, Canada

[2]Dessau LVM, Montreal, Canada

[3]Ecole Polytechnique de Montreal, Department of Civil, Geological and Mining Engineering, Montreal, Canada

*Vincent.Moreau@polymtl.ca

Abstract Life cycle inventories are data intensive by definition and missing data continues to hinder more complete and accurate assessments. This article proposes a statistical approach to address data gaps in life cycle inventories applied to large scale hydroelectric power. The procedure relies on relationships between the technical characteristics of hydropower plants and the material and energy flows necessary throughout the life cycle of such systems. With highly flexible estimators known as kriging, predicting the value of material and energy flows suddenly becomes more accurate. From relatively small sample sizes, kriging allows better estimation without averaging out any of the original data. Similarly, parameter estimation and model validation can be performed through cross validation which assumes very little on the data itself. Mean absolute errors for various forms of kriging and regression show that the former performs better than the latter, more so in cases of incomplete data.

1 Introduction

Hydroelectric power generation accounts for approximately 16% of world electricity output, with China, Canada and Brazil being the leading producers [1]. Despite widespread claims of hydroelectricity being a renewable source of power, very few studies on its potential impact exist. Indeed, less than a handful of life cycle assessments (LCAs) have been conducted on hydro power plants [2]. Hydroelectric projects depend on numerous local conditions, topography, hydrology, geology as well as factors affecting human populations and the environment. This specificity of hydroelectric power accounts for much of the lack of necessary data, making inventory analysis and impact assessments notoriously difficult. Moreover, new hydropower projects became larger over time, multiplying data collection efforts and challenging the development of environmental assessment tools [3]. Generic hydropower plants hardly exist.

Quantifying and qualifying the gains and loses to society and the environment from hydropower projects is a monumental task the World Commission on Dams managed to undertake [4]. Most studies have emphasized the site specific issues which translate into questionable comparisons with other large sources of power. Hence the goal of this article, taking a statistical approach to enable better estimation of life cycle inventory data on hydropower plants and more generally, showing how data gaps in life cycle inventories (LCIs) and associated databases can be alleviated with appropriate statistical estimators. Besides primary sources of data, this article draws upon existing inventories, namely the ecoinvent database and Itaipu assessment [2,5]. A description of the hydropower systems follows before the methodological approach is explained. Results emphasize the importance of the construction phase of hydropower and a conclusion ends this article.

2 System characteristics

The power of hydroelectric plants usually depends on two factors, the water flow and the hydraulic head or height difference between water intake and turbine according to the relation:

$$P = \eta \rho g Q H$$

where P is power (W), η is a coefficient of efficiency, ρ is the density of water (~1000 kg/m3), g is the acceleration due to gravity (~9.8 m/s^2), Q is the flow rate (m^3/s) and H is the head (m). In practice this is implemented with either low head, high flow (generally run-of-river plants), high head, low flow (generally with reservoirs) or combinations in between. Again this scale varies considerably with respect to site specifications.

Overall, hydroelectric power plants are 82 to 88 % efficient [5]. The difference between run-of-river and reservoir plants lies essentially with storage which is one of the key advantages of hydroelectric dams [6]. Electricity cannot be stored such that supply must match demand almost instantaneously. Technically, hydropower plants can respond and adjust their production within minutes and are therefore well suited for both base and peak load production, particularly with a reservoir.

Despite the specificity of every power plant and reservoir, a set of characteristics was chosen based primarily on data availability. Table 1 summarizes the characteristic variables of hydroelectric power plants referred to in this article.

Tab.1: System characteristics

Characteristic variables	Ranges	Notes
Type	0/1	Run-of-river/Reservoir
Total installed capacity (MW)	20 – 14000	-
Annual production (GWh/year)	90 – 90000	Average over ≥ 7 years
Hydraulic head (m)	6 – 1400	-
Surface of reservoir (km^2)	0 – 4200	-
Volume of reservoir (km^3)	0 – 140	-

3 Methodology

Life cycle inventory, as defined by international standards, requires the quantification of material and energy flows as well as emissions crossing a system's boundaries [7]. The quality of a LCI directly influences the overall quality of an LCA [8]. A number of reasons lead us to hypothesize that statistical estimation can overcome both the specificity of hydropower with respect to construction sites and operations and the limitations of current practices dealing with missing data in LCI. Analyzing the inventory of electricity generation tends to yield higher uncertainties when compiled from generic data [9].

Moreover, statistical models have been relatively successful in such cases as with the assessment of chemical manufacturing [10]. Regression and other linear methods have also been applied to inventory analysis of electricity generation [11]. A more flexible estimator is presented here. Borrowed from spatial statistics, kriging distinguishes itself on several accounts. Clearly no cartesian points exist in LCIs such that the characteristic variables described above become coordinates and the material and energy flows or emissions to be estimated, dependent variables. Kriging is a weighted linear combination of the observations. We assume a model with both a random function and a deterministic drift or low order polynomial. Provided a set of unbiasedness constraints are met, the properties in table 2 hold:

Tab.2: Properties of the kriging estimator [12]

Properties	Description	Expression
Unbiased estimator	Expected values of estimates and observations are equal	$E\left[\hat{Z}(x_i)\right]=E\left[Z(x_i)\right]$
Exact estimator	No loss of information	$\hat{Z}(x_i)=Z(x_i)$
Screening effect	Weights vary according to the distance from estimates	-
Size and position	Covariance functions *C(h)* to model observations in space.	$h=\|x_i-x_j\|$
Smoothing effect	Kriging varies less than the estimated phenomena.	$Var(\hat{Z}(x_i))\leq Var(Z(x_i))$

where *E[ ]* and *Var( )* are the expectation and variance operators respectively. The x_i's are characteristic variables and the *Z*'s corresponding material and energy flows. Covariance functions *C(h)* translate in formal terms the idea that distinct observations close to one another should agree more than if they were far apart. A number of valid covariance functions exist, linear, exponential, spherical, etc. Each function has 3 parameters, the first of which is a nugget effect which captures variations on a small scale and can be understood as an interpolating smoothing parameter [13]. Note that since kriging is an exact estimator, it is discontinuous at every observations. The second parameter is a structured variance which equals the total variance when added to the nugget effect. Third, the range of a covariance function corresponds to the distance *h* at which the covariance is nil or observations are unrelated.

If deterministic and random components as well as covariance functions impart great flexibility to the kriging estimator, another advantage is multivariate kriging, or cokriging. Multivariate kriging enables the estimation of primary variables using data from secondary variables simultaneously [14]. All estimations can be performed jointly as the order between primary and secondary variables can be interchanged [15]. For example, if data on explosives is available for the construction of most power plants and dams whereas the excavated volumes are not, joint estimation might provide more accurate results for excavation. In general, primary and secondary variables need not be observed at the same points.

The relatively small sample sizes (here N = 27) available for power generation systems would forfeit many attempts to validate a model and its parameters.

Cross-validation however, has no prior assumption such as normality and observations take part in both trial and validation sets [16]. Leave-one-out cross-validation (LOOCV) successively removes observations and derives estimates from neighboring values exactly once. Moreover, with cokriging, the values of secondary variables for a given observation can be part of the validation data all together or one after the other. These two options are referred to as Leave-one-observation-out cross-validation (LOoOCV) and Leave-one-variable-out cross-validation (LOvOCV), respectively. Comparing the results of the two validation approaches indicate some of the advantages of cokriging, should the errors calculated from LOvOCV be lower than that of LOoOCV.

4 Results

Data on the construction, operation and removal of hydropower plants is scarce. A sample from different sources and for widely different power plants was carefully assembled. Emphasis was put on the construction phase, less controversial than reservoir flooding and responsible for most of the overall impact in many cases [2,5,17]. Time is clearly an important factor to be considered with construction. In particular, Ribeiro and his colleague [2] have shown the LCI of hydroelectric power plants to be sensitive to time horizons. Optimistically the useful lives of infrastructures and equipments were adjusted to 100 and 50 years, respectively. If dams and equipment can technically last that long, chances are operations and economics dictate otherwise.

By and large the main material and energy flows by weight or volume involved in hydroelectric power construction and operation are water, cement and concrete, steel of various grades for structures and equipment, explosives, and fuels and lubricants for machinery, transportation and operation. Their provision and the construction phase itself account for an important share of hydropower's total impact, although water use is often ignored. Because of different scales, the data was normalized, dividing variables by their mean values. Mean absolute errors (MAEs) can then be derived from performing leave-one-observation-out cross-validation (LOoOCV). As shown in figure 1, comparisons between the different estimators are not straightforward.

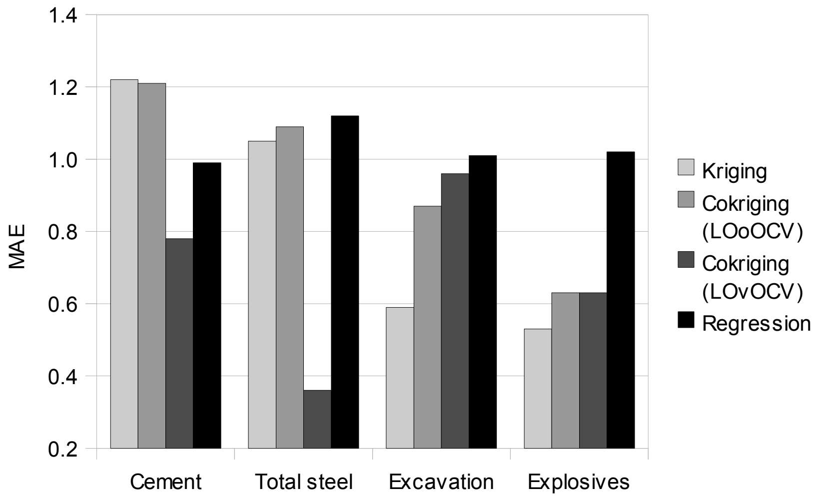

Fig. 1: Comparison of MAEs for different estimators

The most evident result is the most interesting: the kriging and cokriging errors are lower than that of regression for the two rightmost flows of the chart. This shows not only that kriging performs better in the presence of data gaps but also provides more accurate estimates. Estimating one variable at a time reduces somewhat the MAEs, as the comparison of cokriging errors indicates. Both observations support the original hypothesis that estimators as versatile as kriging are particularly well suited in situations where data is missing and needed.

Does accounting for several variables simultaneously further reduce kriging errors? Besides providing more information from which to estimate missing values, the benefits are not obvious either. In the case of kriging, univariate or multivariate, both characteristics, installed capacity and annual production, enter in the calculation. Regression uses installed capacity only. It appears that the advantage of multivariate over univariate kriging (and regression) are particularly important when data is relatively complete, as shown on the left of figure 1. To the contrary, univariate kriging shows lower errors when data is missing. One explanation is weaker relationships between the different material flows themselves than with the characteristic variables, which is a reason why they are characteristic. Hence the preliminary work necessary to identify appropriate variables describing the properties of a system.

5 Conclusion

The lack of data affects inventory analysis of many hydropower plants and other systems. In this article, the authors show how material and energy flows involved in the initial phases of hydropower plants can be better estimated based on their design and operating characteristics. The constraints and drawbacks with respect to data collection and data gaps can therefore be loosened, thanks to appropriate and accurate statistical estimators. The versatility of kriging allows for better estimation especially in the absence of complete data. Limitations to this procedure exist, kriging is better suited for interpolation purposes. A minimum sample size (typically $N = 30 – 50$) is necessary to establish covariance functions, such that this experiment used predefined functions and cross-validation to estimate the better model and its parameters. While no prior assumptions on the data are needed, the results of this parameter selection might differ from one data set to the next. Nevertheless, the approach presented above contributes not only to the estimate of missing data but also to much needed representative data which lacks from existing inventory databases.

6 References

[1] IEA, *Key World Energy Statistics*, International Energy Agency, 2010.

[2] F.D.M. Ribeiro and G.A. da Silva, Life-cycle inventory for hydroelectric generation: a Brazilian case study, *Journal of Cleaner Production*, vol. 18, 2010, pp. 44-54.

[3] D.M. Rosenberg et al., Large-scale impacts of hydroelectric development, *Environmental Reviews*, vol. 5, 1997, pp. 27-54.

[4] World Commission on Dams, *Dams and Development: A new framework for decision making*, London, Sterling VA: Earthscan Publications Ltd., 2000.

[5] C. Bauer et al., Wasserkraft, *Sachbilanzen von Energiesystemen: Grundlagen fur den okologischen Vergleich von Energiesystemen und den Einbezug von Energiesystemen in Okobilanzen fur die Schweiz*, Paul Scherrer Institute Villingen, Swiss center for life cycle inventories Dubendorf: 2007.

[6] D. Egre and J.C. Milewski, The diversity of hydropower projects, *Energy Policy*, vol. 30, 2002, pp. 1225-1230.

[7] ISO, *14040 Environmental management - Life cycle assessment - Principles and framework*, International Organization for Standardization, 2006.

[8] I. Mongelli, S. Suh, and G. Huppes, A structure comparison of two approaches to LCA inventory data, based on the MIET and ETH databases, *International Journal of Life Cycle Assessment*, vol. 10, 2005, pp. 317-324.
[9] J.R. May and D.J. Brennan, Application of data quality assessment methods to an LCA of electricity generation, *International Journal of Life Cycle Assessment*, vol. 8, 2003, pp. 215-225.
[10] G. Wernet et al., Molecular-Structure-Based Models of Chemical Inventories using Neural Networks, *Environmental Science & Technology*, vol. 42, 2008, pp. 6717-6722.
[11] H. Hondo, Life cycle GHG emission analysis of power generation systems: Japanese case, *Energy*, vol. 30, 2005, pp. 2042-2056.
[12] E.H. Isaaks and M.R. Srivastava, *Applied Geostatistics*, Oxford University Press, 1989.
[13] D. Marcotte and M. David, Trend Surface Analysis as a Special Case of IRF-k Kriging, *Mathematical Geology*, vol. 20, 1988, pp. 821-824.
[14] H. Wackernagel, *Multivariate geostatistics : an introduction with applications*, Berlin: Springer, 2003.
[15] D. Marcotte, Cokriging with matlab, *Computers & Geosciences*, vol. 17, 1991, pp. 1265 - 1280.
[16] D. Marcotte, Generalized Cross-Validation for Covariance Model Selection, *Mathematical Geology*, vol. 27, 1995, pp. 659-672.
[17] A. Peisajovich, *Etude du cycle de vie de l'electricite produite et transportee au Quebec*, Direction principale Communication et Environnement, Hydro-Quebec, 1997.