The International Journal of Life Cycle Assessment
https://doi.org/10.1007/s11367-019-01674-y

LCA COMMUNICATION AND LCA FOR ISO LABELS

Extending sensitivity analysis using regression to effectively disseminate life cycle assessment results

Giovanni Di Lullo[1] · Eskinder Gemechu[1] · Abayomi Olufemi Oni[1] · Amit Kumar[1]

Received: 28 June 2018 / Accepted: 13 August 2019

Abstract

Purpose Bottom-up–based life cycle assessment (LCA) approaches are used to assess the greenhouse gas emissions of various products such as transportation fuels. Bottom-up spreadsheet–based models include numerous calculations and assumed values that are uncertain. Currently, most LCAs provide point estimates with a simple one-at-a-time sensitivity analysis, which provides limited insight into how the model assumptions affect the results. Additionally, the LCA models are generally presented with a limited number of scenarios to avoid overwhelming the reader; however, this limits the usefulness of the work, as each reader will be interested in different scenarios. The goal of this work is to use a global sensitivity and regression to provide as much information to the reader as possible in an easily digestible form.

Methods The Morris and Sobol global sensitivity methods are examined to determine if they can accurately identify the key inputs that have the largest effect on overall output variance. A multiparameter linear regression is then used to simplify the model into a single equation. Rstudio and Excel VBA are used to create an easy-to-use template called the Regression, Uncertainty, and Sensitivity Tool (RUST) that can be inserted into any Excel-based LCA model. This method is applied to the previously published FUNdamental ENgineering PrinciplEs-based ModeL for Estimation of GreenHouse Gases in Conventional Crude Oils and Oil Sands (FUNNEL-GHG-CCO/OS) as an example case.

Results and discussion Both the Morris and Sobol methods can identify the key parameters, but the Morris method requires less than 1/100th as many model evaluations. Of the model's 65 parameters, 14 key parameters were identified. The corresponding regression model was found to have an accuracy of ± 0.5 g CO_2 eq/MJ 90% of the time and a maximum error of + 3 and − 1 g CO_2 eq/MJ.

Conclusions This work found that the Morris method can be used to screen key parameters and that a stepwise multiparameter linear regression approach can be used to develop a simplified version of the model. The developed RUST Excel workbook can be used to perform the sensitivity and regression analysis of any Excel-based LCA models. The regression model can then be easily published, it does not require a large effort to make a user friendly version of the model, and it conceals confidential data if necessary. The simplified model makes it easy for policy markets to investigate how changes in critical parameters affect the LCA results without having to learn how to use the full complex model.

Keywords GHG · Interpretation · Life cycle assessment · Morris · Regression · RUST · Sensitivity · Sobol · Uncertainty

1 Introduction

Life cycle assessment (LCA) is used to evaluate environmental performance of a product system taking into consideration the full life cycle stages from cradle-to-grave. According to the ISO framework, LCA has four basic stages: goal and scope definition, life cycle inventory analysis, life cycle impact assessment, and interpretation (ISO 2006a, b). LCA modeling involves multiple assumptions and is subjected to uncertainties because various sources are used. The uncertainties include temporal and spatial variability, input

Responsible editor: Serenella Sala

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11367-019-01674-y) contains supplementary material, which is available to authorized users.

✉ Amit Kumar
Amit.Kumar@ualberta.ca

[1] Department of Mechanical Engineering, 10-263 Donadeo Innovation Centre for Engineering, University of Alberta, Edmonton, Alberta T6G 1H9, Canada

Published online: 26 August 2019

parameters, and model structure uncertainties. Ross et al. (2002) found that only 3% of LCA studies report a quantitative uncertainty analysis. The ISO framework and principles (ISO 2006a, b) recommend performing a sensitivity analysis but provide no guidelines on which methods to use. Sensitivity and uncertainty analyses are important parts of LCA modeling. For example, when a comparative LCA is carried out with the aim of identifying which product has higher emissions, understanding the variations in the results is important. When an attributional LCA is performed, a sensitivity analysis can be used to determine which parameters need to be further investigated to improve model accuracy or which insensitive parameters can be fixed. The sensitivity analysis can also be used to identify areas of improvement that will have the greatest impact on the model output.

This work aims at developing a framework to extend sensitivity analysis by using a regression analysis to create a simplified proxy model, which can be used to help disseminate LCA results. LCA can be performed using a top-down aggregated approach or a bottom-up fundamental calculation approach. The bottom-up LCA examined in this work uses engineering fundamental principles (EFPs) to calculate the energy and mass balances for each process stage, rather than using regional or facility-aggregated energy and material data to approximate emission intensity. Building bottom-up LCA in Excel does not require expensive speciality software; however, there is difficulty in assessing the model sensitivity and uncertainty in a simplistic way. Unlike conventional LCA, the life cycle inventory of EFP LCA commonly relies on complex nonlinear calculations. EFP LCA has been used to compare the GHG emission intensities of different products. The Greenhouse Gases, Regulated Emissions, and Energy Use in Transportation (GREET) model, for instance, compares conventional gasoline and diesel vehicles to alternative battery electric and hydrogen vehicles (Argonne 2015). The Oil Climate Index (OCI) (Gordon et al. 2015), FUNdamental ENgineering PrinciplEs-based ModeL for Estimation of GreenHouse Gases in Conventional Crude Oils and Oil Sands (FUNNEL-GHG-CCO/OS) (Di Lullo et al. 2016; Di Lullo et al. 2017; Nimana et al. 2015a, b, c; Rahman et al. 2014; Rahman et al. 2015), TIAX (Rosenfeld et al. 2009), and Jacobs (Keesom et al. 2012; Keesom et al. 2009) compare the GHG intensities of transportation fuels produced from multiple crude oils. The EFP LCA models in Excel typically examine only a single environmental problem (e.g., GHG emissions, g CO_2 eq), unlike conventional LCA modeled in programs such as SimaPro, which examines numerous impact categories with hundreds of pollutants. As a result, the EFP LCA inventory matrix is smaller and easily manageable in Excel. However, multiimpact EFP models can still be produced in Excel. The EFPs involve rigorous modeling of foreground processes, resulting in nonlinear and strong interaction effects. Additionally, unlike conventional LCA, EFP LCA models do not require matrix calculations. Similar to conventional LCA approaches, EFP LCA models include multiple uncertain inputs; it is therefore important to understand how small changes in the input values can affect their results. A sensitivity analysis involves calculating the partial derivative for each parameter of interest to determine how much the output changes as each input changes over a range. This paper focuses on the application of global sensitivity to these Excel-based EFP LCA models; however, the process can be expanded to conventional LCA that uses existing commercial software.

EFP LCA models focusing on the transportation fuels industry do not perform global sensitivity analysis, however, they do use a simple one-at-a-time (OAT) local sensitivity analysis (Keesom et al. 2012; Keesom et al. 2009; Nimana et al. 2015a, b, c; Rahman et al. 2014; Rahman et al. 2015; Rosenfeld et al. 2009). OAT sensitivity starts with a base case and each parameter is varied separately to calculate the partial derivatives. It is considered a local sensitivity method as the analysis only evaluates the partial derivative around the base case. This provides a limited view of the model's true sensitivity. If the model is nonlinear, the partial derivatives will be dependent on the input parameter's value. If there are interactions between multiple parameters, the partial derivative will depend on the values of the other parameters. More advanced global sensitivity analysis can be used to evaluate the sensitivity across the entire parameter space and account for interaction effects. To illustrate the importance of using a global approach, consider a scenario where a portion of the system's heat is produced using a cogeneration facility, with the remaining heat comes from a convectional boiler. If the base case share of heat from cogeneration were small, the efficiency of the cogeneration facility would appear to be an insensitive parameter (OAT local approach). However, as the share of heat from cogeneration increases, the efficiency parameter's sensitivity would also increase (global approach). Therefore, using a local approach could lead the modeler to ignore or underestimate significant parameters.

Campolongo et al. (2007) illustrated how a Morris sensitivity analysis can be used to accurately identify key parameters in a 60-parameter, nonlinear model. Iooss and Lemaître (2015) provided an overview of various global sensitivity analyses and their applications to a river model. Campolongo and Braddock (1999) used the Morris method to assess an IMAGE greenhouse model. Saltelli et al. (2004) published an entire textbook on global sensitivity in scientific models. Groen et al. (2017) examined the squared standardized regression coefficient, squared Spearman (rank) correlation coefficient, Key Issue Analysis, random balance design, and Sobol local sensitivity methods for use in LCAs. Despite the numerous publications on using global sensitivity analyses, OAT sensitivity analysis is still commonly used. GREET includes a high-level stochastic simulation but provides limited data for input probability distributions and uses small sample sizes. FUNNEL-GHG-CCO/OS includes a

Monte Carlo simulation for uncertainty but still uses OAT sensitivity screening (Di Lullo et al. 2016; Di Lullo et al. 2017).

There have been numerous reviews on the multiple methods available for performing sensitivity and uncertainty in LCA, the most recent from Igos et al. (2019). Exiting sensitivity methods include correlation, key issues, linear regression analysis, marginal, moment-independent modeling, Morris, OAT, scenario, and Sobol/FAST analysis (Ferretti et al. 2016; Groen et al. 2017; Igos et al. 2019). This work focuses on the Morris and Sobol methods as they have an optimal balance of flexibility and ease of use. The Morris and Sobol methods can be applied to nonlinear and nonmonotonic problems with large input variances unlike the correlation analysis, key issues analysis, linear regression, marginal, and OAT methods. The Morris method is also faster than the factored scenario method as fewer samples are required. This is based on simple mathematical theory of the two methods.

While various definitions exist for classifying uncertainty, this work will use Walker et al. (2003) definition, which groups uncertainty under three terms: parameter (empirical values), normative (subjective choices), and model (simplifications and approximations). The analysis purposed here is focused on parameter and normative uncertainties. The process can be expanded to model uncertainty through the exploration of alternative models.

One of the key advantages of OAT local sensitivity analysis is its ease of implementation. Users do not require any statistics or programming background; they can simply use Excel data tables. While there is open source software such as SIMLAB (2016) and PSUADE (Gan et al. 2014) that simplifies the process of performing a global sensitivity analysis, the users still need to automate the process of generating the output from the Excel model. R (2017) and MATLAB (Mathworks 2018) packages are available for both sensitivity and regression analyses but require programming experience and can be difficult to use. Excel add-ins such as ModelRisk by Vose (2019) and CrystalBall by Oracle (2019) are limited to performing Monte Carlo simulations. LCA software such as SimaPro (Pré Consultants B.V. 2019), OpenLCA (GreenDelta 2018), and GaBi (thinkstep 2019) also limited to Monte Carlo simulations only. To overcome the lack non-Monte Carlo sensitivity and uncertainty software, this work created the Regression, Uncertainty and Sensitivity Tool (RUST). While RUST is limited to Excel-based LCA models, the process can be extended to other programs such as SimaPro and OpenLCA which support scripting; however, a software specific interface is needed and is outside the scope of the current work.

Unlike previous work that only focuses on applying sensitivity methods to identify critical parameters, this work takes the analysis one step further by using multiparameter regression to create a simplified version of the complex EFP LCA model. The Excel-based EFP LCA models typically have multiple worksheets with numerous calculations. Given the complexity of the Excel model, the excel workbooks are not always published, and when they are, they are difficult to navigate, especially for users without the required technical background. Additionally, the original modelers can only publish detailed results on a finite number of scenarios, and readers may want to examine scenarios specific to their interests and integrate the results into their own work, which can be difficult without full model results. It also may not be possible to publish the full model due to confidentiality concerns. By using linear regression, a single equation can be created to represent the entire model. The regression equation is easier to publish and can be easily used by nontechnical readers to understand the response of the model. The regression equation can also be integrated into other models, such as the integrated resource planning models (e.g., Long-range Energy Alternatives Planning System (LEAP) (Heaps 2016)), extending its usability. The regression equation can be calculated at significantly higher speeds than the full model, allowing more detailed assessments such as hybrid Monte Carlo methods (Pedroni and Zio 2012). Currently, if researchers want to incorporate results from a published LCA into their work, they would only be able to use the discrete values provided with the publication. For example, if practitioners were to perform an LCA of gasoline from bitumen and want to use results from a study on the transportation emissions of bitumen via pipeline or rail, they would be forced to select the emission factor provided by the paper even if it does not align with their current model. This is also useful for inputs that may change with time. For example, older LCA study results were needed to be recalculated when the Intergovernmental Panel on Climate Change (IPCC) global warming potentials (GWP) were updated. However, if a regression model was provided based on key inputs such as electricity emission factor, pipeline diameter, GWP time frame, etc., the formula could be integrated into the gasoline LCA and adjusted to match the specific scenario required in the larger model. The gasoline LCA would still require a sensitivity and uncertainty analysis of its own, which may determine that some of the inputs in the regression model are no longer sensitive and can be fixed.

Assume the pipeline model was insensitive to an input X1 (such as the "diesel emission factor"), and it was therefore not included in the regression model. If the larger gasoline LCA model is sensitive to X1, it would be better to include the input in the pipeline regression model. However, the error from ignoring X1in the pipeline model will be relatively small. If the gasoline LCA model is sensitive to X1, the contributions will primarily come from other stages of the gasoline model, not the pipeline regression model. While a fully rigorous

model would be more accurate, the difficulties in publication constrained its availability.

Overall, the main goal of this work is to create a general framework to help LCA practitioners effectively communicate their results and maximize the usability of their models. This goal will be accomplished in two parts: (1) by using global sensitivity analysis to communicate which parameters are important and how model assumptions can affect the results and (2) by using regression analysis to create a simplified version of the model, whose results are easy to publish. This framework is aimed at modelers who work primarily with spreadsheets. An Excel-based template that can be integrated into any existing Excel model is created. The coding for the template is done using VBA and R languages. While there are multiple global sensitivity methods, this work focuses on the Morris and Sobol methods. The Morris method requires minimal computational effort, and the Sobol method showed the best overall performance in the study by Groen et al. (2017). The FUNNEL-GHG-CCO/OS model is used as an example case. It assesses the GHG intensity of transportation fuels, gasoline, diesel, and jet fuel produced from different crude oils. For completeness, RUST does include the ability to perform a Monte Carlo uncertainty analysis, but such feature is not discussed in this work.

2 Methods

This work has two main tasks: to perform a global sensitivity analysis to identify the critical parameters and to create a regression model using the critical parameters (Fig. 1). Before the critical parameters can be identified, the Morris and Sobol methods must be validated to determine how many model evaluations are required to accurately identify the critical parameters. Linear regression is then used on the critical inputs to create an approximate model. The regression equation is then validated and checked for overfitting. To accomplish these tasks, an Excel template, RUST, was created to link the R scripts in a user-friendly format. Multiple R packages are used within the template (Carnell 2016; Carnell 2017; Grothendieck 2013; Hadley Wickham 2017; Koenker et al. 2017; Pujol et al. 2017; R Core Team and contributors worldwide 2017; Regis Pouillot and Denis 2017; Wickham and Chang 2016). The FUNNEL-GHG-CCO/OS model was used to illustrate how the tool can be used (Di Lullo et al. 2016; Di Lullo et al. 2017). The RUST template can be used with any Excel-based LCA model and supports parallel computing to reduce computational time, see the RUST demonstration for further information (Electronic Supplementary Material—Online Resource 1). The approach could also be extended to exiting LCA software by creating application specific interfaces, but this aspect is beyond the scope of this paper.

2.1 Global sensitivity methods

A basic OAT local sensitivity analysis evaluates the model at the minimum and maximum values for each parameter, a process that requires only $N = 2 \times k + 1$ model evaluations, where k is the number of parameters. The OAT method is commonly used as it is quick and easy to implement, but it provides no information on the interactions between the inputs. To overcome this limitation, the developed model allows evaluating the Morris expanded OAT method and the Sobol variance method, which examine the primary and interaction effects (global methods). As mentioned in Sect. 1, Morris and Sobol were selected as they provide the best balance of flexibility and ease of use. Key issues analysis (KIA) was not investigated as it is limited to models with small uncertainty ranges as it assumes a linear model response (coefficient of variation < 5%) (Groen et al. 2017).

2.1.1 Morris method

Sensitivity is a partial derivative that can be dependent on the value of the parameter in question (nonlinear) or the values of the other parameters (interactions). The Morris method uses a design of experiments approach to evaluate the partial derivative across the parameter space (Morris 1991). The user selects the number of levels, p, and number of OAT designs/approaches, r. The number of model evaluations required is $N = r \times (k + 1)$, where k is the number of parameters. Values for p and r should be even and r should be larger than p to ensure uniform sampling of the space (Morris 1991). Typical r values are between 10 and 50 (Campolongo et al. 2007). The level specifies the number of unique values each parameter can take. In Fig. 2, $p = 4$, meaning each model parameter (X1, X2, X3) can take on a value of 0, 0.33, 0.67, or 1. The values are then scaled-based on the input ranges provided. Figure 2 shows the sample path (blue line) in which a starting point is randomly selected. One input is changed at a time, from points 1–4 and 5–8, to calculate the partial derivatives of the respective parameters. From points 4 and 5, multiple parameters are changed simultaneously, which results in a new approach. The number of approaches, r, also sets the number of derivatives calculated for each parameter. Higher r values are required to ensure that the parameter space is adequately sampled. For example, Fig. 2 shows that no samples were taken in the bottom corner. The red line represents an example model response and illustrates the effects of the number of levels. Since only 4 levels are used, the calculated partial derivatives will always be negative. However, a positive slope between X3 = 0.85 and 1 would be detected if more levels were used. The derivative in the X1 direction would incorrectly appear to be negative over the entire domain. For nonlinear equations, large p values are required to ensure that variability in the sensitivity is accurately captured. Campolongo et al.

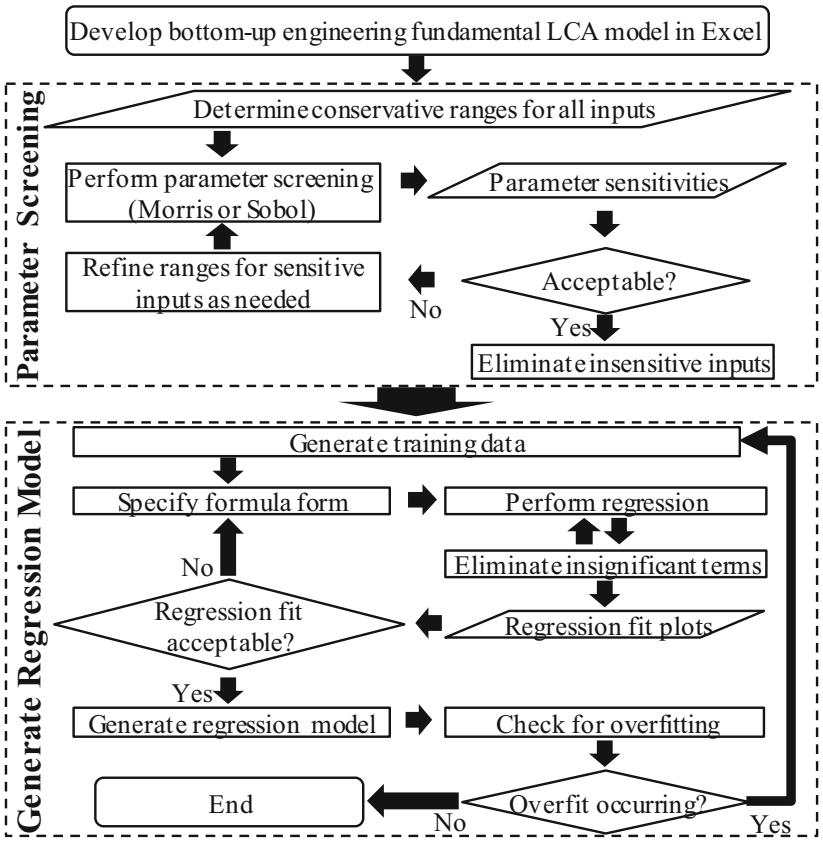

Fig. 1 Overview of general process

introduced an optimization approach that started with 500–1000 *r* approaches and then determined which 10 approaches provided the best spread of the parameter domain (Campolongo et al. 2007). When the model's evaluation time is long, Campolongo's method can reduce the number of model evaluations; however, the brute force approach used to select the optimal approaches can be computationally expensive. Since the FUNNEL-GHG-CCO/OS model can calculate over 100 samples, it is quicker to simply assess a larger number of approaches than to use Campolongo's optimization method.

The partial derivative, also known as the elementry effect, for parameter *j* is determined for each OAT design, $i = 1$ to r, using Eq. (1):

$$D_j^i = \frac{f\left(X^i + \Delta\right) - f\left(X^i\right)}{\Delta} \tag{1}$$

where X^i is input ($i = 1, \ldots, k$) and $f(X^i)$ is the model output with the respective input. Delta is determined from the number of levels, p, as $\Delta = s/(p-1)$. The step size is generally set to $s = p/2$; in Fig. 2, s = 1 (Pujol et al. 2017). To simplify the presentation of the results, the mean (μ) of the absolute partial derivatives and the standard deviation (σ) are calculated from Eq. (2) and Eq. (3), respectively:

$$\mu_j = \frac{1}{r}\sum_{i=1}^{r} |D_j^i| \tag{2}$$

$$\sigma_j = \sqrt{\frac{1}{r}\sum_{i=1}^{r}\left(D_j^i - \frac{1}{r}\sum_{i=1}^{r} D_j^i\right)^2} \tag{3}$$

The larger the μ, the more sensitive the model output is to the input value. The standard deviation (σ) shows how the sensitivity changes throughout the parameter domain. If σ is large, then either the partial derivative is affected by the input value as it moves across its range (nonlinear) and/or the slope depends on the values of the other parameters (interactions). The absolute value of the partial derivative is taken for the mean to ensure nonmonotonic parameters are not incorrectly labeled as insignificant; the standard deviation does not use absolute values as doing so would underestimate the true variation in a nonmonotonic model. The mean and standard deviation values are normalized using the parameter ranges so that they are comparable. The Morris method groups the inputs into three main categories: those with negligible effects (small μ and σ), linear effects with negligible interactions (large μ and small σ), and nonlinear/large interaction effects (large σ).

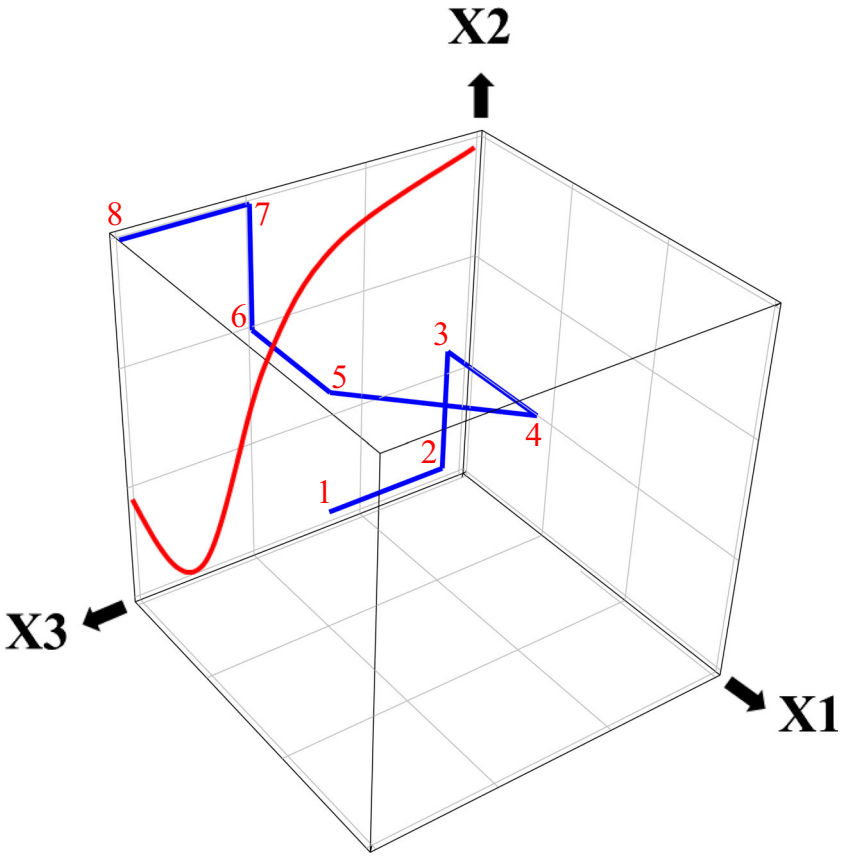

Fig. 2 Example of Morris sampling with three parameters (k), two OAT designs (r), and four levels (p)

2.1.2 Sobol indices

The Morris method identifies whether parameter sensitivities vary across the domain space; however, it neither can differentiate between nonlinear and interaction effects nor can it identify which specific parameters are interacting. The Sobol method uses a decomposition of variance approach to identify what percentage of the overall variance can be attributed to each parameter or group of parameters (Gan et al. 2014; Iooss and Lemaître 2015; Pujol et al. 2017; Saltelli et al. 2004; Sobol 1993).

An ANOVA analysis is used to compare the variation between groups to the variation within the groups and identify if the means of the groups are significantly different. The first-order Sobol indices calculate the main effect of each variable, while the Sobol total indices calculate the contribution to variance due to the main and all interaction effects. Figure 3 shows how two $n \times k$ matrices can be used to generate k AB_i matrices with column i from A and the remaining columns from B. The model is then run for all the samples in A, B, and AB_i for a total of $N = n \times (k + 2)$ model evaluations, where n is the number of samples for each matrix (rows) and k is the number of inputs (columns).

Matrix A [*f(A)*]

	X_1	$X_{...}$	X_k
1	0.3	0.2	0.1
2	0.4	0.5	1.0
....	0.1	0.7	0.2
n	0.6	0.8	0.5

Matrix B [*f(B)*]

	X_1	$X_{...}$	X_k
1	0.6	0.7	0.3
2	0.5	0.4	0.6
....	0.1	0.2	0.9
n	1.0	0.6	0.8

Matrix AB_1 [*f(AB 1)*]

	X_1	$X_{...}$	X_k
1	0.6	0.2	0.1
2	0.5	0.5	1.0
....	0.1	0.7	0.2
n	1.0	0.8	0.5

Matrix AB_1 [*f(AB...)*]

	X_1	$X_{...}$	X_k
1	0.3	0.7	0.1
2	0.4	0.4	1.0
....	0.1	0.2	0.2
n	0.6	0.6	0.5

Matrix AB_1 [*f(AB k)*]

	X_1	$X_{...}$	X_k
1	0.3	0.2	0.3
2	0.4	0.5	0.6
....	0.1	0.7	0.9
n	0.6	0.8	0.8

Fig. 3 Sobol sample generation

The 1st and total Sobol indices for each parameter, i, are then calculated as follows:

$$S_i = \frac{\frac{1}{n}\sum_{j=1}^{n} f(B)_j * \left(f(AB_i)_j - f(A)_j\right)}{\mathrm{Var}(Y)} \quad (4)$$

$$S_{Ti} = \frac{\frac{1}{2n}\sum_{j=1}^{n} \left(f(AB_i)_j - f(A)_j\right)^2}{\mathrm{Var}(Y)} \quad (5)$$

In layman's terms, first-order indicia take the variation in the results that occur from changing the value of the i^{th} parameter ($f(AB_i)_j - f(A)_j$), scaling the results using ($f(B)_j*$) and normalizing the effect against the total variance using /Var(Y). The sum of all primary and interaction terms equals 1; hence, in a purely additive model, the sum of the first-order indices will equal 1. Parameters with low total Sobol indices can be removed, as their effect on the overall variance is insignificant and will have minimal effect on the output.

2.1.3 Assessing Morris and Sobol accuracy

Since the Morris method uses a random number generator to determine the sample points, results vary between runs. To assess the accuracy and stability of the Morris method, each scenario was run 50 times, and the average and standard deviation of the Morris μ and σ are reported. A high accuracy means the Morris method correctly identifies the sensitive parameters, and a high stability implies that the results will not drastically change between runs. The various scenarios used different r (4, 10, 20, 40, 60, 80, 100) and p (4, 8, 16, 32) values to examine their effects on the Morris plot accuracy and stability. Even p values are used to ensure equal probability sampling for each level (Campolongo and Cariboni 2007). To ensure all p levels are sampled, the number of approaches r should be larger than p. The number of scenarios used was determined iteratively based on the model's response in an effort to minimize the number of scenarios required. Minimal changes in stability where observed from $r = 80$ to 100; therefore, higher r values were not examined. Since p produced a weaker response than the r value, fewer levels were selected.

Sobol analysis uses $n = 2 \times 2^x$ samples, where x varies from 3 to 15; x is selected by the user and increases the samples required by a factor 2 as the error in the Sobol indecencies is $O(n^2)$. The number of model executions required, $N = n \times (k + 2)$, where k is the number of parameters, can quickly become computationally expensive. To assess the effect of sample size on accuracy, x was varied from 3 to 10 for all 65 parameters (full model) and from 3 to 15 for 27 parameters (reduced model). The reduced model was created by removing the parameters that had a negligible effect on the output to reduce processing time; the Morris method and preliminary results from the Sobol method were used to reduce the number of parameters. While Sobol indices are positive, and the sum of first-order indices should be less than or equal to 1 by definition, negative indices and sums greater than 1 can occur if the sample size is too small due to numerical error. Bootstrapping is used within the R code to determine the standard error for each index.

To compare the Morris and Sobol methods, the parameter ranks were used. The Morris method used μ, and the Sobol method used the total indices to rank the parameters. A rank of 1 is given to the parameter with the highest sensitivity.

2.2 Multiparameter linear regression model

Linear regression is used to create a simplified version of the final model. The least squares regression approach calculates coefficients for each term and an intercept value by minimizing the squared residual error. The p values are used to determine which terms are significant and identify the "best" model.

2.2.1 Identifying critical parameters

LCA models can have numerous inputs; however, uncertainty is generally dominated by a small percentage of key inputs, which can be identified using a sensitivity analysis. Due to the size and complexity of modern models, it is not feasible to perform an in-depth sensitivity analysis on all the inputs. Screening methods are used to quickly eliminate insignificant inputs. A sensitivity analysis can also aid modelers in identifying errors within the model by identifying unexpected responses.

Care should be taken to use input-specific ranges for the sensitivity analysis rather than a generic ± % for each input, since the output sensitivity is dependent on the input's probable range. For example, in a pipeline model, relative roughness can vary by multiple orders of magnitude, while pump efficiency will vary by only ± 20%. If a generic range is used, the relative roughness may incorrectly appear to be insignificant. If possible, multiple inputs should be combined. For example, global pump efficiency should be used rather than multiple individual pump efficiencies that all have the same range, as the global efficiency will provide a more conservative estimate of the sensitivity. Since the regression model, which will be developed in the next step, is only valid within the parameter ranges used to create the training dataset, narrow ranges, which limit the regression model's scope, should be avoided. In the first stage, conservatively large ranges are used for simplicity and to ensure that no sensitive parameters are prematurely removed. Additional time is then invested to refine the parameters' range for the sensitive parameters. Figure 1 provides a high-level overview of the sensitivity screening process. The Sobol analysis is computationally

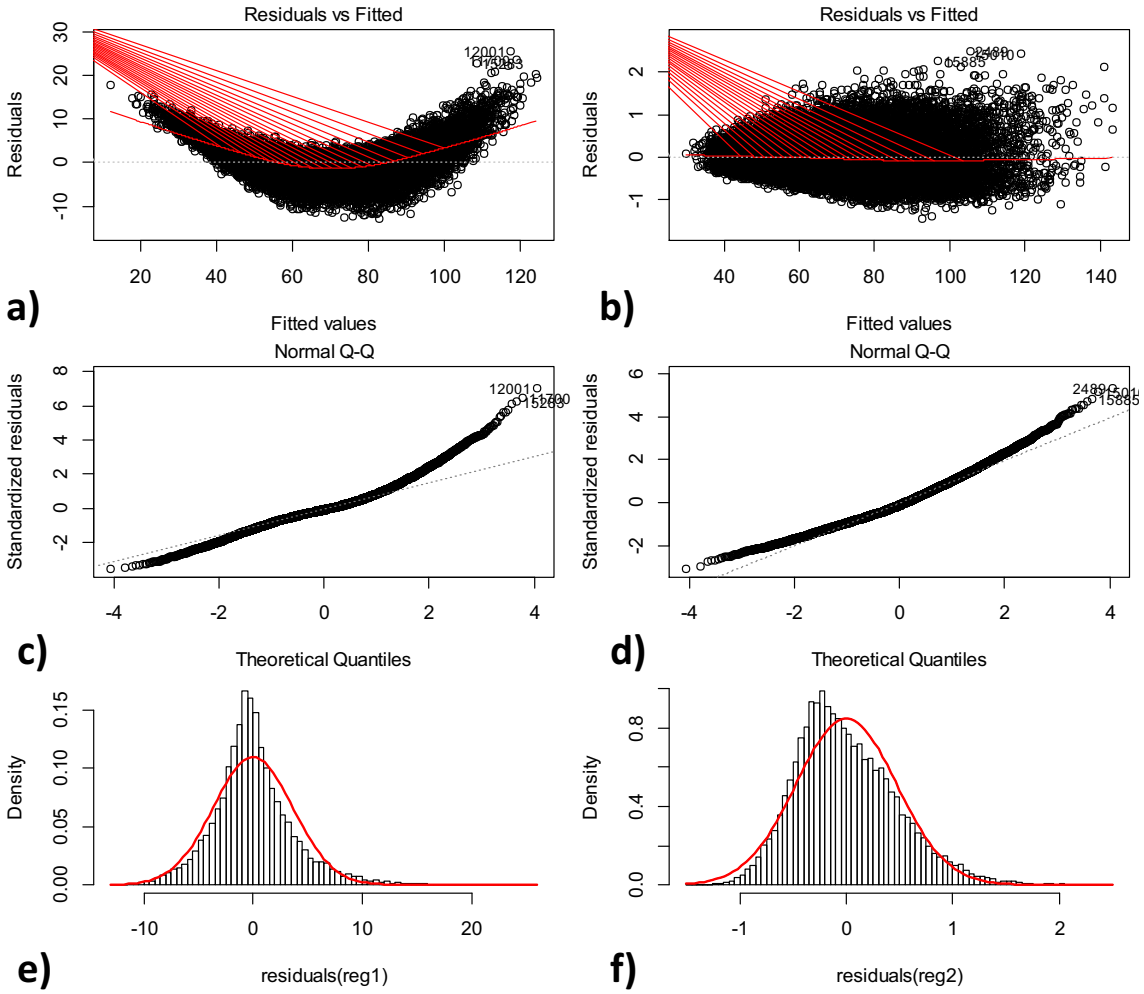

Fig. 4 Regression verification example with a poor fit (left) and a better fit (right) showing residuals vs. fitted (**a** and **b**), normal Q-Q plots (**c** and **d**), and residal histograms (**e** and **f**)

intensive but can be used to provide further insight into the interactions between parameters; however, the Morris method was found to be sufficient for identifying the critical parameters.

2.2.2 Developing the regression model

The training data is generated using Latin hypercube sampling (LHS) in the open source software R, rather than conventional Monte Carlo sampling. In Monte Carlo sampling, each point is generated independently using a random number generator. Relying on a random number generator is inefficient and can result in points being either closely clustered together or spread far apart. LHS attempts to generate samples that are evenly spread across the entire parameter domain. Each input is split into N intervals, and a sample is taken from each interval. As a result, LHS sampling error is O(1/N), significantly lower than Monte Carlo O(1/sqrt(N)) (Chrisman 2014). However, LHS advantage only applies to single distributions. When multiple-parameter distributions are used, the LHS method ensures the samples for each parameter are evenly distributed but does nothing to ensure an even sampling of parameter combinations (Chrisman 2014). Therefore, LHS advantage is reduced as the number of *significant* parameters increases. While models can have over a hundred inputs, the output uncertainty is generally dominated by only a few inputs. Therefore, LHS can still be advantageous even in large models (Chrisman 2014).

Uniform distributions should be used to ensure that the entire sample space is uniformly sampled. If triangle or normal distributions are used, there would be fewer training samples at the edge of the parameter space, which would lead to larger errors when extreme values are used. At least one sample is required for each term; however, a general rule of thumb is to use 10–15 samples per term to avoid overfitting (Frost 2017). More samples may be needed if there is a high degree of collinearity. The regression model includes terms for each parameter as well as all of their interactions. Interaction terms are made by multiplying multiple parameters to determine their combined affect. The highest level of interaction is specified and includes all the lower interaction levels as well. For example, third-level interactions include every possible two-

and three-term combinations of the parameters. The total number of terms (Nt) in the regression model is determined from the sum of the combination terms, that is, $Nt = \sum_1^L \frac{p!}{L!(p-L)!}$, where p is the number of parameters and L is the level of interaction. The higher levels of interactions generally lead to a lower residual error but require additional computational time and larger samples.

The goal in using a regression model is to find which combination of parameter and interaction terms best replicates the underlying model. There are multiple approaches to model selection, but none is perfect, and expert judgment is required (Frost 2012; Frost 2015; Olejnik et al. 2000). Models with a large number of parameters and levels of interaction can contain thousands of terms, resulting in billions of possible regression model combinations. Since it is not possible to evaluate every combination, a stepwise approach is used instead of a best subset approach (Frost 2012; Frost 2015). The process starts by including every term and iteratively eliminating the terms with the highest p values. When a term's p value is less than 0.05, there is less than a 5% chance that this term is irrelevant and that the coefficient is zero; hence, terms with larger p values are eliminated, as their coefficients are most likely 0. The two-sided p values are calculated within R using the t statistic which is calculated from the standard errors of the coefficients. Multicollinearity between terms leads to significant instability in the p values, which makes it difficult to determine which terms to eliminate (Frost 2013). Since the interaction terms will be correlated with the individual parameter terms, multicollinearity will occur even if the parameters in the original model are not correlated (X1 correlated with X1 × X2 terms). By centering the data before running the regression analysis, the multicollinearity of the inputs is significantly reduced (Frost 2013). Once a term is eliminated and the regression analysis is updated, the p values for the remaining terms will be updated and if there is multicollinearity within the data, a term that appears to be insignificant may become significant. As a result, terms are removed in an iterative fashion, rather than simultaneously, until only terms with p values less than 2×10^6 remain.

To increase the accuracy of the regression model, nonlinear squared and cubed terms may be required. In order to avoid multicollinearity between the primary (X1), squared ($X1^2$), and cubed ($X1^3$) terms, an orthogonal polynomial was required. An orthogonal polynomial uses a weighting function to ensure the dot product over a specified interval is equal to 0. This transformation is accomplished using the polynom library in R (Venables et al. 2016). To minimize the regression determinant matrix size, squared and cubed terms should only be included for inputs that are suspected to have nonlinear effects. In order to identify potential nonlinear effects, an OAT sensitivity analysis was performed using 5 samples for each input. Excel's built-in trendline function was then used on each input's 5 samples to identify nonlinear trends. Further information on this process is available in Online Resource 1 (Electronic Supplementary Material). Morris cannot be used to identify nonlinear inputs as it cannot differentiate between nonlinear and interaction effects.

2.2.3 Validating regression model

A valid regression model should have a high adjusted R^2 value and a normally distributed residual error across the entire output domain. Low R^2 values suggest a nonlinear model may be needed. Alternatively, by using a higher level of interaction can improve the R^2 value. To ensure that the errors are normally distributed, residual vs. fitted plots, normal Q–Q plots, and residual histograms are used. The residuals vs. fitted plots can be used to determine if the residuals are uniformly distributed along the output domain (Chouldechova 2017). The red line in Fig. 4 (residuals vs. fitted) is a smoothed average of the residuals; ideally, it should be flat and equal to zero. In Fig. 4, the curved residuals indicate a common issue where the regression model is attempting a linear fit to a nonlinear model. However, the conning effect (residual increases as fitted value increases) can indicate multiple nonlinear effects are not being accounted for (Fig. 4b). If the residuals are normally distributed, then the normal Q–Q plots should produce a straight diagonal line (Fig. 4c, d) (Ihaka 2007). An upward curve on the right or a downward curve on the left of the residual normal Q–Q plots indicates a heavier than normal tail. Alternatively, a downward curve on the right or an upward curve on the left of the residual normal Q–Q plots indicates a lighter than normal tail.

The histograms can be used to illustrate the Q–Q plots; the red line is an ideal normal distribution for comparison (Fig. 4e, f). Ideally, the model residuals should match the red line, with the median at zero. A nonzero median suggests a bias in the model. Skewed distributions with a long tail suggest the regression model is inaccurate within a portion of the parameter domain and may indicate the insufficient coverage of the parameter domain within the training sample.

2.2.4 Developing and validating the regression model

The regression model is developed using the method shown in Fig. 1. To avoid errors in the p values, the data is centered to reduce multicollinearity, as discussed in Sect. 2.2.1. Due to the complexity of the model, a stepwise approach is used to determine the final regression model form. Given the large number of terms, it is not practical to eliminate only one parameter at a time. Instead, all the terms with a p value above a specified test point are removed. To avoid removing a large number of parameters at once, the test point starts at 0.9 and progressively decreases to 2.5e−16.

Once the iterative calculation is completed, a text file is loaded into Excel with a summary of the output. The output includes the minimum, 5th, 25th, 75th, and 95th percentile, and maximum residual values; the number of terms in the regression model; the Akaike information criterion (AIC); and the mean squared error (MSE) for each iteration. The Akaike information criterion balances the goodness of fit and the number of terms in the regression model. The lower the AIC value, the more efficient the regression model is. The user can select which model form to use by balancing trade-offs in accuracy and number of terms. The residual vs. actual plots, normal *Q*–*Q* plots, and residual histograms are generated for the optimal model to verify that the residuals are normally distributed.

Once the regression coefficients have been determined, it is important to check for overfitting. The regression model should replicate the true model; however, when overfitting occurs, the regression model begins to fit random patterns within the training data and falsely suggests an accurate regression model has been determined. If overfitting occurs, using new samples could result in significantly larger residuals. A general rule of thumb is to use 10–15 samples per term to avoid overfitting (Frost 2015; Frost 2017); however, more samples may be needed if there is a high degree of collinearity. The overfitting macro generates multiple samples using unique seeds and compares the average and standard deviation of the residual minimum, 5th percentile, 95th percentile, and maximum values. Large standard deviations indicate overfitting is most likely occurring. To prevent overfitting, the number of samples used to train the regression model should be increased.

2.3 Model template

R Studio is used to perform both the sensitivity and regression analyses. While R Studio can be used directly to perform the analyses, a template was created to aid those to make it user friendly. The users simply need to insert their existing Excel model into the RUST file and fill in the specified inputs. Excel macros programmed in VBA will prepare the data and execute Rscripts to perform the analysis. The main inputs are folder locations for saving .txt and .csv files, high-level inputs for the various functions, and the parameter table. The parameter table specifies the value ranges and provides the input cell address for each parameter. The Excel macros will then generate the required sample file, run the samples through the user's model, generate an output file, perform the desired analysis, and output the results (Fig. 5). A "mapping inputs" macro is provided to streamline the process. A demonstration the process is provided (Online Resource 1—Electronic Supplementary Material). The Morris and Sobol analyses were run using the Pujol et al. (2017) R library.

2.4 Case study

The FUNNEL-GHG-CCO/OS model was used as an example in this study; full details of the LCA are available in (Di Lullo et al. 2016; Di Lullo et al. 2017). The bottom-up EFP LCA model calculates the GHG intensity of gasoline produced from various crudes. This case study focuses on Maya, Bow River, and Athabasca mined bitumen. Maya is a heavy Mexican crude produced off the coast of Mexico using offshore drilling and nitrogen injection, Bow River is a heavy convectional crude produced using water flooding in southern Alberta, and Athabasca bitumen is produced in northern Alberta using open surface mining. Each crude has 44–65 inputs. The life cycle GHG emissions were accounted for each cured per the functional unit of g CO_2 eq/MJ of gasoline. The scope of the LCA includes drilling the well, crude production, surface processing, crude transportation, refining, transportation fuel distribution, and combustion stages. Construction phase was not included, due to their negligible contribution to total emissions. The energy input requirements at each life cycle stage were examined. The main pieces of equipment included were boilers/heaters, pumps, and compressors. The energy requirements for each equipment and the associated GHG emissions were calculated using technical parameters such as unit efficiencies, production ratios, and operating temperatures and pressures. A full list of inputs examined can be found in the Online Resource 2 (Electronic Supplementary Material). While the case study uses only 65 parameters, RUST is capable of assessing larger input models as the 64-bit version of Excel is no longer limited to 2 GB of RAM (Microsoft Support 2017). Furthermore, the case study uses a single output, but multiple outputs can be specified in RUST for the Morris and Sobol analysis. Regression must be run one output at-a-time or with a weighted output.

3 Results

Section 3.1 focuses on the global sensitivity analysis; the accuracy for the Morris and Sobol methods is determined and compared to identify which method is best. Section 3.2 focuses on the regression analysis; the number of terms required and levels of interaction required are determined. The regression model is then validated and checked for overfitting.

3.1 Global sensitivity validation

The Morris and Sobol methods are examined to determine if they can identify the critical inputs. The two methods are then compared.

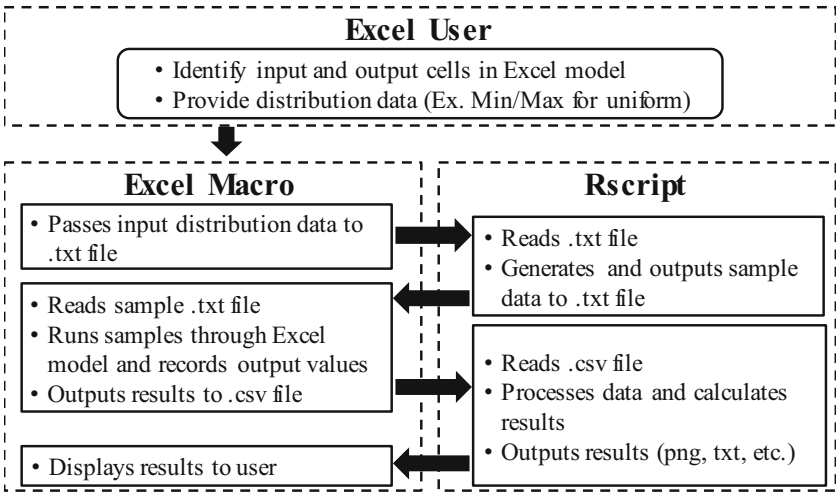

Fig. 5 RUST background process flow diagram

3.1.1 Validating Morris

To validate the Morris method, 50 runs are averaged for various r and p values. The number of runs used is arbitrary; 50 gives an adequate balance between accuracy and computing time. Two parameters were examined (shown in Fig. 6): the gas-to-oil ratio (GOR) and the pipeline velocity for crude transportation to the refinery. For the Maya pathway, the GOR was the most significant parameter; it had the largest μ and σ on the Morris plot. The pipeline velocity parameter was examined as it results in a nonlinear response. The standard deviation in the Morris μ and σ values over the 50 trials were examined; only the μ results are shown in Fig. 6. If the standard deviation is large, the Morris plot results will change each time the analysis is run and can lead to incorrect representations of the model sensitivity. The variability in the GOR and pipeline velocity parameters' Morris μ and σ decreased as the number of OAT designs, r, increased from 4 to 40. However, negligible benefit is seen in using r values greater than 40. The number of levels, p, can reduce the variance, especially when small r values are used. When there are many levels, more OAT designs are required to thoroughly sample each level and prevent variations between runs.

The σ standard deviation followed a similar pattern. The μ and σ values averaged over the 50 trials were all within 5% of each other, indicating that they all converge to the same result if enough trials are used. Therefore, if enough samples are used to reduce the standard deviation, the Morris plot will converge to the correct solution and only one run is required. Overall, an r value of 20–40 will provide sufficient sampling to minimize the variability in the Morris plot results.

3.1.2 Validating Sobol

The standard error is calculated by the Sobol package for each index using a bootstrapping approach. The results in Fig. S1 (Electronic Supplementary Material) are for the 27-parameter scenario; results were similar for the 65-parameter scenario. The standard error is approximately the same for all 27 first-order indices regardless of the index's value. The standard error for the total indices is dependent on each index's value; the normalized average across the 27 parameters is shown in Fig. S1 (Electronic Supplementary Material). Overall, the error in the total indices is lower than the error in the 1st-order indices.

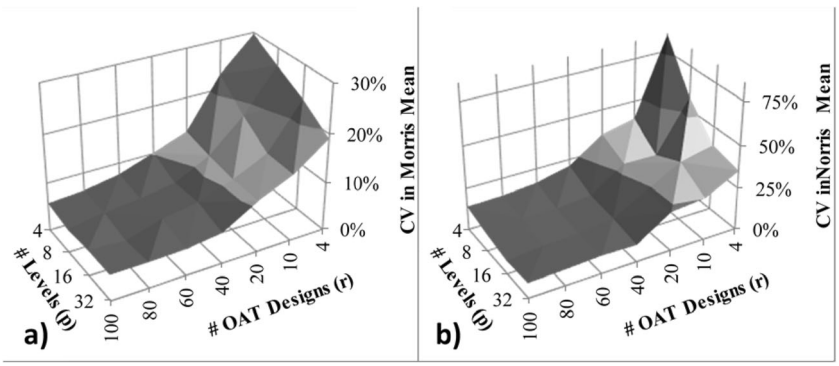

Fig. 6 Morris μ coefficient of variation for the GOR (**a**) and pipeline velocity (**b**). The average GOR and pipeline velocity Morris μ are 13 and 0.5 g CO_2 eq/MJ, respectively

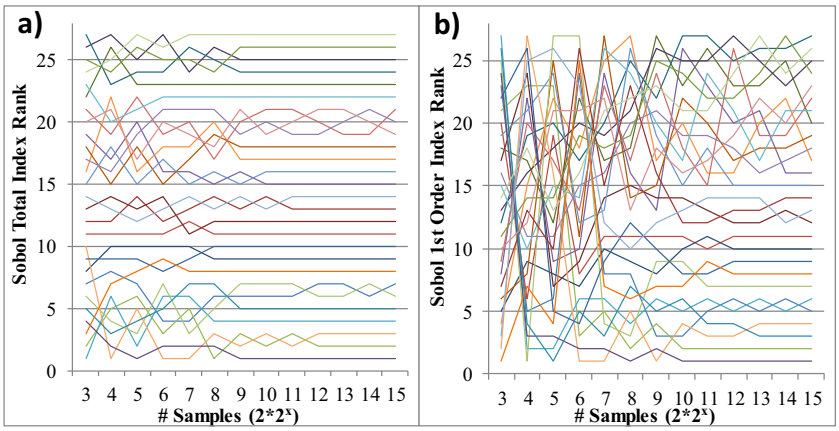

Fig. 7 Variation in total (**a**) and 1st-order (**b**) Sobol index ranks and values vs. sample size (each line represents a parameter; 27 parameters were used)

Input ranks based on their 1st and total Sobol index are shown in Fig. 7; a flatline indicates that the ranks are not changing as the number of samples used is increased. The Sobol total indices' rank stabilizes once x is above 8. Sobol first-order indices had large standard deviations, and their values were unstable and had negative indices, until x was greater than 12; the ranks of the less important inputs did not stabilize. The accuracy and stability of the second-order indices are similar to the first-order indices. In conclusion, a large number of samples are needed to produce accurate Sobol indices. If only the total indices are needed to help screen out insensitive inputs, fewer samples can be used.

3.1.3 Morris vs. Sobol accuracy

To compare the Sobol and Morris methods, we examined the ranks for the 65- and 27-parameter scenarios. The Morris method used $r = 60$ and $p = 32$, the Sobol method used $k = 13$, and the OAT method used 2 samples per input plus the base case. The Morris method required 3960 and 1680 model evaluations, and the Sobol method required 548,864 and 237,568 model evaluations for the 65- and 27-parameter scenarios, respectively. The parameter ranking between the OAT method and Morris can vary by up to 21 positions (Fig. 8a). Unlike the Sobol methods, even the ranking of the most significant variables is inaccurate using the OAT method. While the ranks vary by up to seven positions for the Sobol method (see Fig. 8b), the ranks of the significant parameters are more stable; hence, both the Morris and Sobol methods can accurately identify the significant parameters. However, the Morris method requires less than 1/100th the number of model evaluations and is ideal for screening purposes. Once insensitive parameters have been removed by the Morris method, the Sobol method can be used to identify which parameters are interacting, if desired.

3.2 Multiparameter linear regression

Users of the template would not need to perform the Morris and Sobol verifications but would instead start with parameter screening. The Morris method was used with 60 OAT designs (r) and 32 levels (p), and the results were used to produce Fig. 9. Each data point represents a parameter. The Excel template generates the Morris plot and adds data labels, so it is easy to identify each parameter. The data labels were removed in Fig. 9 to reduce the image size. The parameters in red box A were eliminated, as they were insensitive and did not have significant nonlinear or interaction effects. It is important to look at the actual value of the Morris μ and σ when deciding when to eliminate parameters rather than just relative position as an excessively large parameter can skew the axis scale. Selecting the cutoff point between sensitive and insensitive is subjective. The axis scale should be considered, as well as the clustering of parameters. For example, a parameter can be deemed insensitive if the Morris mean is less than 1% of the base case value and has a low Morris standard deviation. For the Fig. 9 Morris plot, the inputs in box B have similar sensitivities; hence, they should either all be included or excluded as a group. The parameters in the top right with $\mu \approx 6$–8 and $\sigma \approx 4$–5 g CO_2 eq/MJ illustrate the advantage of

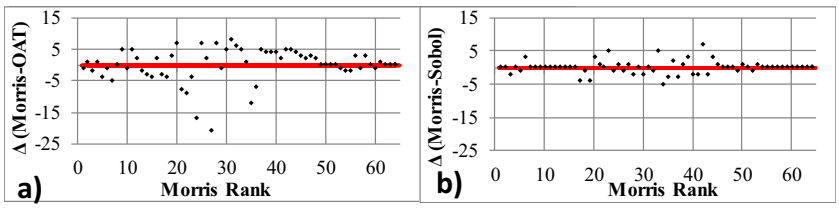

Fig. 8 Variations in parameter rank: OAT (**a**) and Sobol (**b**) (plus means the OAT/Sobol rank's parameter as less important than the Morris method)

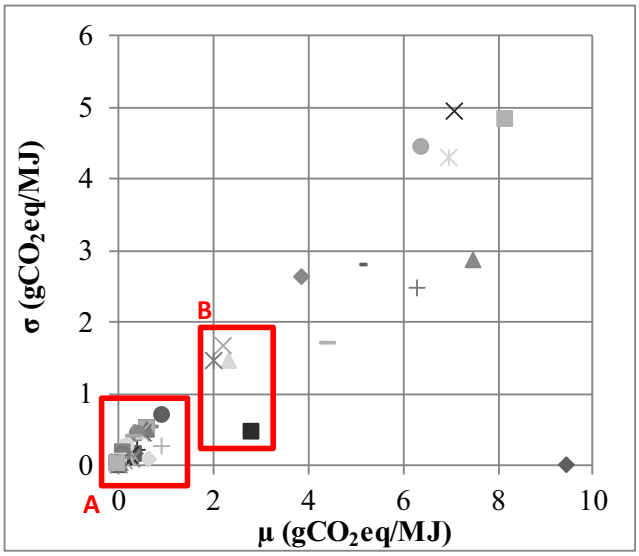

Fig. 9 Morris screening of FUNNEL-GHG-CCO/OS Maya pathway

the Morris method. Depending on the base case used in an OAT local sensitivity analysis, these parameters could have sensitivities of 1–2 g CO_2 eq/MJ and be grouped with the insensitive parameters. The Sobol method was not required in this scenario as the Morris method sufficiently reduced the number of parameters.

3.2.1 Validating the regression model form

The regression analysis was performed for the Maya pathway using the 14 critical parameters with 1st, 2nd, 3rd, 4th, and 5th level interactions. Three scenarios with training sample sizes of 3000, 5000, and 10,000 samples are used for comparison. For 5 levels of interaction, a minimum of 3472 samples is required; hence, 5th level interactions are not included in the 3000-sample scenario. Since an iterative method is used to narrow down which terms are included in the final regression model, the number of samples per term will increase as terms are eliminated, thereby reducing the chance of overfitting in the final model. Theoretically, if the number of samples is too small, then potentially significant terms could be eliminated during the initial iterations. However, the results show that the number of samples used does not have a detectable effect on the accuracy of the final form of the regression model. The accuracy was strongly related to the level of interaction and the number of terms included (see Fig. 10); this relationship will be model-specific. Caution should be used when assessing the accuracy of the regression model versus the number of terms included, as overfitting can occur. As shown in Fig. 10, moving from 320 to 316 terms causes a jump from 2.05 to 2.27 g CO_2 eq/MJ in the residual variance (max residual–min residual). An overfit check with 100 runs of 1000 samples each on both the 320 and the 316 term models found that the residual variance ranges were 1.8–3.2 and 3.0–3.2 g CO_2/MJ, respectively, indicating they both have the same level of accuracy when tested against new samples. The residual variance uses the maximum and minimum residuals to ensure no outliers occur. While it may be tempting to use a regression model with more terms to improve accuracy, the model should be verified with a new sample to ensure the improved accuracy is not due to overfitting.

3.2.2 High-level checks

For the Maya pathway, the 5000-sample, 4th-level interaction regression model was used. The predicted vs. actual, residual vs. actual, normal *Q*–*Q*, and residual histogram plots are generated for the final regression model with 84 terms (Fig. S2—Electronic Supplementary Material). The actual vs. predicted plot shows no large deviations/outliers; however, there is an indication that the regression model will under predict emissions when actual emissions are large. The residual vs. predicted plot confirms that the regression model under predicts emissions for the extreme scenarios and that there is a slight bias in the residuals. The normal *Q*–*Q* plots' upward curve on the right indicates that the right tail is slightly longer than for a normal distribution, which is illustrated by the histogram. Overall, the plots indicate that the residual is approximately normally distributed with no major problems, but it should be noted that the residual error will increase as the predicted value increases; this conning effect may indicate nonlinear effects are not being accounted for.

3.2.3 Overfitting checks

To check for overfitting, the range of residual variance (max residual–min residual) across 100 runs, each with 1000 samples, is compared to the training set residual variance (Table 1). When 4th level interactions are used with 14

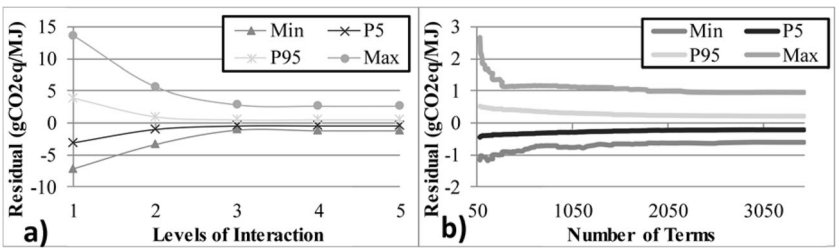

Fig. 10 Effect of levels of interaction used (**a**) and number of terms included in the final regression model (**b**) on the residual (both use 5000 samples in training data; right plot uses 5th level interactions)

Table 1 Range of residual variance (max–min) with 4th level interactions, ranges from 100 runs with 1000 samples each (the same overfit check samples were used in every scenario)

Scenario	All terms		Final form	
	Overfit check	Training set	Overfit check	Training set
3000	36,590–114,890	1.72	2.41–4.86	4.07
5000	13,670–53,290	1.95	1.94–4.08	3.84
10,000	13,180–38,250	2.49	1.91–3.58	3.09

parameters, the total number of terms is 1470; when all terms are included in the regression model, the number of samples per term is 2, 3.4, and 6.8 for the 3000, 5000, and 10,000 sample scenarios, respectively, all of which are lower than the suggested 10–15 samples per term. As a result, the residual variances from the overfit check are significantly larger than the training set variances. The final form of the regression model includes only 46, 43, and 39 coefficients, which result in 77, 116, and 217 samples per term for the 3000-, 5000-, and 10,000-sample scenarios, respectively; hence, their residual variance is similar to the training set and indicates that overfitting has not occurred (Table 1).

3.2.4 Checking model response

A well-fitted model should be able to accurately predict both the output values and sensitivities (partial derivative). The accuracy of the output values has already been addressed. To determine the accuracy of the partial derivative, the Morris method is used. If the regression model perfectly predicts the emission intensity, the μ and σ values of the partial derivative should be the same, providing the same sample is used. Overall, the regression model accurately captures the model response. The average and standard deviation values of the derivative are accurate to within 4% and 15%, respectively (Table S1—Electronic Supplementary Material). The large error in the Morris σ for the refinery emissions is acceptable since the value is approximately zero. The regression model under predicts the σ in the crude low heating value (LHV) correction factor while maintaining accuracy in the μ, suggesting that there may be errors in the regression model's predictions for the extreme scenarios.

3.3 Final regression models

Once the regression formula has been determined, an Excel macro generates the final formula from the text file. Excel formulas are limited to 8192 characters; hence, long formulas may need to be split between two or more cells. Overall, the original FUNNEL-GHG-CCO/OS model, which included hundreds of calculations and multiple Excel worksheets, can be simplified down to the 14-parameter regression model. The same approach was applied to the Bow River water flood crude, and Athabasca mined bitumen crudes from Alberta, Canada; the regression accuracies are shown in Table 2. Each model contains between 58–96 terms and 1655–2989 characters, well within Excels 8192-character limit. The regression model inputs include technology flows, emission flows, and characterization factors as shown in the Electronic Supplementary Material (Online Resource 2).

4 Discussion

The results of this work highlight that the Morris method is ideal for parameter screening. It requires only minimum and maximum values for each input; as a result, it can accurately account for interaction and nonlinear effects only like the other methods available using small sample size. Since the Sobol method demands 100× as many samples as the Morris method, hence, it is not recommended for screening purposes. In order to be more widely used, these methods need to be more transparent in the discussion, especially to transfer the message to the less experience practitioners. However, after the insensitive inputs have been removed by the Morris method, the Sobol method can be used to identify which specific interactions are occurring using second-order indices. Results from the OAT analysis showed that ignoring interaction effects can lead to an incorrect estimation of parameter importance. The inaccuracy of the OAT method will vary between models and depends on the extent of interaction and degree of nonlinearity. While the FUNNEL model is nonlinear, it is monotonic which benefits the OAT approach. While Groen et al. (2017) suggest a linear regression model (without interaction terms) can be used for identifying important parameters, it requires many samples, and the coefficients can only be used as importance measures if the models R^2 is close to one indicating a good fit.

When using the Morris method, 20–40 approaches (r) provided accurate results, with minimal benefits above 40 approaches. Parameter ranks varied by an average of ten positions between runs when only 4 approaches (r) were used. One parameter varied by 49 positions between runs; therefore, care should be taken when using a small number of runs. When 10 approaches (r) were used the average rank change fell to 5.5, with a maximum change in rank between runs of 14. The number of Morris samples needed will also depend on the complexity of the model; nonlinear and highly interactive models will require more samples, while linear additive models will require fewer samples. Sobol requires 2×2^8 for screening purposes using the total indices. If the practitioner wants to determine if interaction effects are occurring, then 2×2^{12} samples are needed. Determining second-order indices to identify specific interactions requires nearly doubling the sample size.

Table 2 Regression summary for Maya nitrogen injection, Bow River water flood, and Athabasca mined bitumen

Residual (g CO_2 eq/MJ)	Min	5th Percentile	95th Percentile	Max	No. of terms	No. of characters
Maya	− 1.16	− 0.44	0.53	2.68	58	1655
Maya (nonlinear)	− 0.63	− 0.08	0.07	0.44	164	4137
Bow River	− 1.17	− 0.54	0.7	2.46	72	2192
Bow River (nonlinear)	− 0.38	− 0.10	0.10	0.55	135	4991
Athabasca	− 0.19	− 0.12	0.17	0.34	96	2989

While the Morris and Sobol methods are restricted to uncorrelated inputs, they are still useful for screening purposes. Correlated inputs can be grouped and represented as a single output, while this method cannot determine the importance of the individual inputs and requires the correlation to be approximated as a direct link, it can identify if the grouped parameter is significant or not. Alternatively, moment-independent methods such as entropy-based sensitivity (Iooss and Lemaître 2015) and PAWN indices are available (Pianosi and Wagener 2015). The Sobol method can also be extended to account for correlations with two additional indices (Rabitz 2010). Correlations are not required in determining the regression model. If the model is accurate when trained with uncorrelated data, it will also be accurate when correlated inputs are used. Since background processes are not always rigorously modeled in EFP LCA, correlations between background emission factors should be investigated. For example, the background emission factors of natural gas, diesel, and fuel oil may all depend on the emission factor of the electrical grid which is utilized by all three processes (Suh and Qin 2017). By grouping the parameters as described above, the potential impact of dependency can be investigated. Data analysis techniques such as principle component analysis (PCA) and partial east squares regression (PLS) are useful when the dataset available is correlated or has a limited sample size; however, since RUST generates an uncorrelated sample to train the regression model these methods are not needed. Furthermore, since PLS emphasizes developing a predictive model, it is not useful as an alternative to the Morris methods for screening out insignificant parameters (Minitab Inc. 2017).

The Maya and Bow river regression models exhibited conning of the residuals which may indicate nonlinear effects that are not being accurately accounted for. Testing the models for nonlinear inputs using the trendline approach described in the Online Resource 1 (Electronic Supplementary Material) indicated that there were multiple nonlinear inputs. However, none of the R^2 values equaled one indicating a more complex response was occurring (Online Resource 2). To test if adding nonlinear terms to the regression model would improve the fit, the Bow river and Maya crudes were rerun and since their residuals were larger than the Athabasca pathway. Maya included a squared term for the crude LHV and compressor driver efficiency, and Bow included crude LHV, pipeline capacity, and pipeline velocity. The results in Table 2 indicate that adding the nonlinear terms reduced the residuals by 40–80% but increased the size of the formula by 65–90%. The nonlinear models did not exhibit any signs of overfitting. Therefore, including nonlinear terms can improve the regression model's accuracy, at the expense of requiring longer computing times and producing a larger regression model.

Unlike the previous works which simply examined the sensitivity methods available (Groen et al. 2017; Igos et al. 2019), this work goes one step further by adding a multiparameter regression model. While publishing the full LCA model is often preferable, it is not always possible due to difficulties making the model user friendly, nondisclosure agreements, or other confidentiality concerns. Furthermore, in some cases, a simplified version of the LCA model is preferable. For example, nontechnical policy makers may want to examine how the LCA results change when key inputs are adjusted without having to learn a complex model. Since journal papers need to be kept concise, researchers are often limited in how many scenarios/alternatives they can present. By including a regression model, the LCA usefulness to fellow researchers and policy makers is increased as they can then examine scenarios specific to their needs. The regression model can also improve the life of a paper by allowing users to update key values such as methane's GWP which is updated each time a new Assessment Report is release by the IPCC (Myhre et al. 2013).

RUST is also capable of running a Monte Carlo simulation but is outside the scope of this study. Currently, RUST supports lognormal, normal, project evaluation and review technique (PERT), modified PERT, triangle, and uniform distributions and uses Latin hypercube sampling. Unlike other Excel add-ins, it does support parallel computing. Output uncertainty is a product of input uncertainty and input sensitivity; therefore, probability distributions are only required for the sensitive inputs. The Morris global sensitivity method can be used to identify sensitive inputs, and a Monte Carlo simulation can then be run using probability distributions for the sensitive inputs only, reducing data collection efforts. The uncertainty simulation can also be run using the regression model rather

than the full model to reduce computational times for complex models. However, running uncertainty on the regression model introduces additional model uncertainty based on the accuracy of the regression model. In the case of the Maya, Bow, and Athabasca crude pathways evaluated in this work, the parameter uncertainty will be an order of magnitude larger than the regression model uncertainty which would make the modeling error negligible. Since the regression model can be evaluated quickly increasingly complex analysis such as hybrid and nested Monte Carlo methods which attempt to separate aleatory (random) and epistemic (known with poor precision) uncertainty and require a large number of model, evaluations can now be performed in less time (Pedroni and Zio 2012).

RUST includes a mapping macro, which can be used to easily run sensitivity and uncertainty on any Excel based LCA model. Furthermore, the Morris and Sobol analysis is performed with scripts, which can be integrated into other programs such as SimaPro and OpenLCA in the future. RUST can also be used with Excel models that are linked to speciality engineering software such as Aspen Hysys.

5 Conclusions

The goal of this paper was to create a framework to help LCA practitioners effectively present their results and maximize the usability of their model. Sensitivity analysis is an important part of LCA; however, it is generally over simplified. This work showed how the Morris and Sobol methods can be used to identify key model parameters by accounting for interaction and nonlinear effects, unlike the one at-a-time method. The Morris method was found to require less than 1/100th as many model evaluations as the Sobol method, making it an effective screening method. The developed RUST tool allows practitioners to easily perform Morris and Sobol sensitivity on their Excel-based LCA models.

It was determined that 20–40 samples per input (approaches) are required to accurately use the Morris method for screening purposes. Using Morris screening, the number of inputs was reduced from between 60 and 70 to 14 for Maya and Bow and 16 for Athabasca. Using the sensitive inputs identified from the Morris method, this work then takes the sensitivity analysis one step further and produces a proxy regression model. The levels of interaction and number of terms were varied until a model with the desired accuracy is obtained.

The regression models were found to have an accuracies of the following: $\text{Maya}^{+2.7}_{-1.2}$, $\text{Bow}^{+2.5}_{-1.2}$, and $\text{Athabasca}^{+0.3}_{-0.2}$ g CO_2 eq/MJ. To improve the fit for the Maya and Bow river pathways, nonlinear terms were added, resulting in new accuracies of the following: $\text{Maya}^{+0.4}_{-0.6}$, and $\text{Bow}^{+0.6}_{-0.4}$. The regression formulas are small enough to fit into a single Excel cell, making the model easy to be published. The regression models can then be used by researchers and policy makers to calculate results for their specific scenarios. It also allows the LCA practitioners to share a simplified version of their model without having to invest large amounts of time in making their complex model user friendly. Furthermore, the regression model can help protect confidential data which would prevent the publication of the full model.

Acknowledgments The authors thank Astrid Blodgett for editorial assistance.

Funding information NSERC/Cenovus/Alberta Innovates Associate Industrial Research Chair in Energy and Environmental Systems Engineering and the Cenovus Energy Endowed Chair in Environmental Engineering provided financial support for this project. Support was also given by the Natural Sciences and Engineering Research Council of Canada (NSERC) Postgraduate Scholarships - Doctoral Program. As a part of the University of Alberta's Future Energy Systems (FES) research initiative, this research was made possible in part thanks to funding from the Canada First Research Excellence Fund (CFREF).

References

Argonne (2015) GREET1. Argonne National Laboratory. https://greet.es.anl.gov/. Accessed June 14, 2016.

Campolongo F, Braddock R (1999) Sensitivity analysis of the IMAGE Greenhouse model. Environ Model Softw 14:275–282

Campolongo F, Cariboni J (2007) Sensitivity analysis: how to detect important factors in large models. Office for Official Publications of the European Communities, Luxembourg

Campolongo F, Cariboni J, Saltelli A (2007) An effective screening design for sensitivity analysis of large models. Environ Model Softw 22:1509–1518

Carnell R (2016) lhs: Latin hypercube samples. https://cran.r-project.org/package=lhs. Accessed Dec. 14, 2017.

Carnell R (2017) triangle: provides the standard distribution functions for the triangle distribution. https://cran.r-project.org/package=triangle. Accessed Dec. 14, 2017.

Chouldechova A (2017) Regression diagnostic plots. Carnegie Mellon University. https://www.andrew.cmu.edu/user/achoulde/94842/homework/regression_diagnostics.html. Accessed Sept. 13, 2017.

Chrisman L (2014) Latin hypercube vs. Monte Carlo sampling. Lumina. http://www.lumina.com/blog/latin-hypercube-vs.-monte-carlo-sampling. Accessed July 13, 2017.

Di Lullo G, Zhang H, Kumar A (2016) Evaluation of uncertainty in the well-to-tank and combustion greenhouse gas emissions of various transportation fuels. Appl Energy 184:413–426

Di Lullo G, Zhang H, Kumar A (2017) Uncertainty in well-to-tank with combustion greenhouse gas emissions of transportation fuels derived from North American crudes. Energy 128:475–486

Ferretti F, Saltelli A, Tarantola S (2016) Trends in sensitivity analysis practice in the last decade. Sci Total Environ 568:666–670

Frost J (2012) Regression smackdown: stepwise versus best subsets! The Minitab Blog. http://blog.minitab.com/blog/adventures-in-statistics-2/regression-smackdown-stepwise-versus-best-subsets. Accessed Sept. 13, 2017.

Frost J (2013) What are the effects of multicollinearity and when can I ignore them? The Minitab Blog. http://blog.minitab.com/blog/adventures-in-statistics-2/what-are-the-effects-of-multicollinearity-and-when-can-i-ignore-them. Accessed May 2, 2013.

Frost J (2015) The danger of overfitting regression models. The Minitab Blog. http://blog.minitab.com/blog/adventures-in-statistics-2/the-danger-of-overfitting-regression-models. Accessed Sept. 3, 2015.

Frost J (2017) Overfitting regression models: problems, detection, and avoidance. Statistics by Jim. http://statisticsbyjim.com/regression/overfitting-regression-models/. Accessed May 26, 2017.

Gan Y et al (2014) A comprehensive evaluation of various sensitivity analysis methods: A case study with a hydrological model. Environ Model Softw 51:269–285

Gordon D, Brandt A, Bergerson J, Jonathan K (2015) Know your oil: creating a global oil-climate index. Carnegie Endowment for International Peace. http://carnegieendowment.org/files/know_your_oil.pdf. Accessed Feb. 18, 2016.

GreenDelta (2018) OpenLCA.

Groen EA, Bokkers EAM, Heijungs R, de Boer IJM (2017) Methods for global sensitivity analysis in life cycle assessment. Int J Life Cycle Assess 22:1125–1137

Grothendieck G (2013) nls2: non-linear regression with brute force. https://cran.r-project.org/package=nls2. Accessed Dec. 14, 2017.

Hadley Wickham JH, Romain Francois, Jukka Jylänki, Mikkel Jørgensen (2017) readr: read rectangular text data. https://cran.r-project.org/package=readr. Accessed Dec. 14, 2017.

Heaps CG (2016) Long-range Energy Alternatives Planning (LEAP) system. Stockholm Environment Institute. https://www.energycommunity.org. Accessed June 14, 2018.

Igos E, Benetto E, Meyer R, Baustert P, Othoniel B (2019) How to treat uncertainties in life cycle assessment studies? Int J Life Cycle Assess 24:794–807

Ihaka R (2007) Quantiles and quantile based plots. The University of Auckland. https://www.stat.auckland.ac.nz/~ihaka/787/lectures-quantiles.pdf. Accessed Sept. 13, 2017.

Iooss B, Lemaître P (2015) A review on global sensitivity analysis methods. In: Dellino G, Meloni C (eds) Uncertainty management in simulation-optimization of complex systems: algorithms and applications. Springer US, Boston, MA, pp. 101–122.

ISO (2006a) ISO 14040:2006: environmental management - life cycle assessment - principles and framework.

ISO (2006b) ISO 14044:2006: environmental management - life cycle assessment - requirements and guidelines.

Keesom B, Blieszner J, Unnasch S (2012) EU pathway study: life cycle assessment of crude oils in a European context. Jacobs Consultancy. http://www.energy.alberta.ca/Oil/pdfs/OSPathwayStudyEUjacobsRept2012.pdf. Accessed June 14, 2016.

Keesom W, Unnasch S, Moretta J (2009) Life cycle assessment comparison of North American and imported crudes prepared for Alberta Energy Research Institute. Jacobs Consultancy. http://www.eipa.alberta.ca/media/39640/life%20cycle%20analysis%20jacobs%20final%20report.pdf. Accessed Feb. 18, 2016.

Koenker R, Portnoy S, Ng PT, Zeileis A, Grosjean P, Ripley BD (2017) quantreg: quantile regression. https://cran.r-project.org/package=quantreg. Accessed Dec. 14, 2017.

Mathworks (2018) MATLAB. Mathworks. https://www.mathworks.com/products/matlab.html. Accessed Feb. 14, 2019.

Microsoft Support (2017) Memory usage in the 32-bit edition of Excel 2013 and 2016. https://support.microsoft.com/en-ca/help/3066990/memory-usage-in-the-32-bit-edition-of-excel-2013-and-2016. Accessed Feb. 2, 2019.

Minitab Inc. (2017) What is partial least squares regression? Minitab Inc., . https://support.minitab.com/en-us/minitab/18/help-and-how-to/modeling-statistics/regression/supporting-topics/partial-least-squares-regression/what-is-partial-least-squares-regression/. Accessed Feb. 14, 2019.

Morris MD (1991) Factorial sampling plans for preliminary computational experiments. Technometrics 33:161–174

Myhre G et al (2013) Anthropogenic and natural radiative forcing. Cambridge University Press, Cambridge and New York

Nimana B, Canter C, Kumar A (2015a) Energy consumption and greenhouse gas emissions in the recovery and extraction of crude bitumen from Canada's oil sands. Appl Energy 143:189–199

Nimana B, Canter C, Kumar A (2015b) Energy consumption and greenhouse gas emissions in upgrading and refining of Canada's oil sands products. Energy 83:65–79

Nimana B, Canter C, Kumar A (2015c) Life cycle assessment of greenhouse gas emissions from Canada's oil sands-derived transportation fuels. Energy 88:544–554

Olejnik S, Mills J, Keselman H (2000) Using Wherry's adjusted R2 and Mallow's Cp for model selection from all possible regressions. J Exp Educ 68:365–380

Oracle (2019) Crystal ball. https://www.oracle.com/ca-en/applications/crystalball/. Accessed Feb. 14, 2019.

Pedroni N, Zio E (2012) Empirical comparison of methods for the hierarchical propagation of hybrid uncertaintly in risk assessment, in presence of dependencies. Int J Uncertainty Fuzziness Knowledge Based Syst 20:509–557

Pianosi F, Wagener T (2015) A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. Environ Model Softw 67:1–11

Pré Consultants BV (2019) SimaPro. https://simapro.com/. Accessed Feb. 14, 2019.

Pujol G et al. (2017) sensitivity: global sensitivity analysis of model outputs. https://cran.r-project.org/package=sensitivity. Accessed Dec. 14, 2017.

R Core Team and contributors worldwide (2017) The R Stats Package. https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html. Accessed Dec. 14, 2017.

Rabitz H (2010) Global sensitivity analysis for systems with independent and/or correlated inputs. Procedia Soc Behav Sci 2:7587–7589

Rahman MM, Canter C, Kumar A (2014) Greenhouse gas emissions from recovery of various North American conventional crudes. Energy 74:607–617

Rahman MM, Canter C, Kumar A (2015) Well-to-wheel life cycle assessment of transportation fuels derived from different North American conventional crudes. Appl Energy 156:159–173

Regis Pouillot M-LD-M, Jean-Baptiste Denis (2017) mc2d: tools for two-dimensional Monte-Carlo simulations. https://cran.r-project.org/package=mc2d. Accessed Dec. 14, 2017.

Rosenfeld J, Pont J, Law K, Hirshfeld D, Kolb J (2009) Comparison of North American and imported crude oil lifecycle GHG emissions - final report prepared for Alberta Energy Reseach Institute. TIAX LLC. http://eipa.alberta.ca/media/39643/life%20cycle%20analysis%20tiax%20final%20report.pdf. Accessed Mar. 25, 2016.

Ross S, Evans D, Webber M (2002) How LCA studies deal with uncertainty. Int J Life Cycle Assess 7:47

Saltelli A, Tarantola S, Campolongo F, Ratto M (2004) Sensitivity analysis in practice: a guide to assessing scientific models. John Wiley & Sons Ltd., West Sussex, England

SIMLAB (2016) SIMLAB. European Union Science HUB. https://ec.europa.eu/jrc/en/samo/simlab. Accessed Feb. 14, 2019.

Sobol IM (1993) Sensitivity estimates for nonlinear mathematical models. Math Model Comput Exp 1:407–414

Suh S, Qin Y (2017) Pre-calculated LCIs with uncertainties revisited. Int J Life Cycle Assess 22:827–831

thinkstep (2019) GaBi Product Sustainability Software. http://www.gabi-software.com/canada/software/gabi-software/. Accessed Feb. 14, 2019.

Venables B, Hornik K, Maechler M (2016) polynom: a collection of functions to implement a class for univariate polynomial

manipulations. https://CRAN.R-project.org/package=polynom. Accessed Feb. 14, 2019.
Vose D (2019) ModelRisk Vose Software. https://www.vosesoftware.com/products/modelrisk/. Accessed Feb. 14, 2019.
Walker WE, Harremoës P, Rotmans J, van der Sluijs JP, van Asselt MBA, Janssen P, Krayer von Krauss MP (2003) Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support. Integr Assess 4:5–17
Wickham H, Chang W (2016) ggplot2: create elegant data visualisations using the grammar of graphics. https://cran.r-project.org/package=ggplot2. Accessed Dec. 14, 2017.