Sustainable Cities and Society 44 (2019) 70–84

Contents lists available at ScienceDirect

Sustainable Cities and Society

journal homepage: www.elsevier.com/locate/scs

Improving life cycle-based exploration methods by coupling sensitivity analysis and metamodels

Sandrine Duprez[a], Marine Fouquet[a], Quentin Herreros[a], Thomas Jusselme[a,b,*]

[a] COMBO Solutions, Lyon, France
[b] Building 2050 Research Group, Ecole Polytechnique Fédérale de Lausanne (EPFL), Fribourg, Switzerland

ARTICLE INFO

Keywords:
Sensitivity analysis
Metamodels
Life cycle assessment (LCA)
Environmental impact
Early design stage

ABSTRACT

Exploration methods combine parametric energy assessments and data visualization to support building designers at early design stages. When exploration methods come to Life-Cycle Assessment (LCA) and the Global Warming Potential (GWP) assessment, a larger number of input parameters induces a very high computation load. Previous researches suggested using Sensitivity Analysis (SA) to decrease the space exploration thanks to their sampling techniques, and input sensitivities. However, this theoretical framework has almost never been applied to building LCA so far and underline two major issues. Upon SA techniques, which one is most suitable for LCA input specificities? How is it possible to extend the exploration process outside the limits of SA samples? This article addressed these questions thanks to an extensive state-of-the-art, the description of a new method combining Sobol SA and Artificial Neural Network (ANN), and a case study. The Sobol method delivered satisfying results with the computation of quantitative indices. Then, an Artificial Neural Network trained on the data generated by the SA was used to predict the GWP of new design alternatives in a small amount of time, and with a coefficient of determination higher than 0.9. Finally, the proposed method adapted exploration methods to the LCA complexity.

1. Introduction

New regulations are targeting the reduction of greenhouse gas emissions to face climate change. The built environment is particularly affected, since it is a major contributor to these emissions. For example, in France, the built environment accounts for 25% of greenhouse gas emissions (Bâtiment à Énergie Positive & Réduction Carbone » Contexte, 2017). As a result, the next French regulation, that will be mandatory in 2020 (Réflexion Bâtiment Responsable 2020-2050 | Informez-vous et contribuez sur une vision prospective du bâtiment responsable à l'horizon 2020-2050, 2018), will include lifecycle targets to limit the Global Warming Potential (GWP). Also, new incentives aiming at reducing the environmental impact of buildings, such as the French state label E + C- have already been introduced (MEEM, 2016).

Thus, it is now essential to take the GWP of a building into account at the early design stage, when the most important decisions are taken. Indeed, it has been pointed out by several studies that taking the right decisions at the early design stage enables to significantly reduce the GWP (Gervásio, Santos, Martins, & Simões da Silva, 2014). Recently, exploration methods have been developed to support designers in making such decisions (Østergård, Jensen, & Maagaard, 2017; Jusselme, Rey, & Andersen, 2018). First, design parameters as well as their range values are defined. In this paper, these varying design parameters will be referred to as "inputs". Then the sampling process of a sensitivity analysis (SA) generates some combinations of inputs, called "design alternatives", and the performance of each design alternative is assessed (output). Finally, data visualization techniques are used to explore the solution space, specifically on design parameters with high sensitivity indices. Exploration methods enable to determine several design alternatives that meet the performance goal, and leave more choices to the designer.

So far, these methods have mostly been applied on the operating energy consumption of buildings, but almost never on the entire building lifecycle. Studying the entire building lifecycle involves new

Abbreviations: LCA, life cycle assessment; GWP, global warming potential; FAST, fourier amplitude sensitivity testing; MLR, multiple linear regression; SVR, support vector machine; ANN, artificial neural network; GA, genetic algorithm; GP, Gaussian process; SA, sensitivity analysis; SRC, standardized regression coefficient; SRRC, standardized rank regression coefficients; PCC, partial correlation coefficient; PRCC, partial rank correlation coefficient; RMSE, root mean square error; ReLU, rectified linear unit; WWR, window to wall ratio; PV, photovoltaic

* Corresponding author.
E-mail address: Thomas.jusselme@epfl.ch (T. Jusselme).

https://doi.org/10.1016/j.scs.2018.09.032
Received 23 May 2018; Received in revised form 27 July 2018; Accepted 25 September 2018
Available online 28 September 2018

inputs that can be chosen by the designer, so the number of dimensions is increased up to at least 15. With such a large amount of dimensions, the exploration of the solutions space is more difficult to achieve. Even with discrete values (e.g. 4 values per dimensions), the resulting sample would exceed a billion of alternatives, inducing a too high computation load. However, some studies have made a first adaptation at applying exploration methods on building LCA (Basbagill, Flager, Lepech, & Fischer, 2013; Heeren et al., 2015; Jusselme et al., 2016). They used SA to sample the input space and get a first insight into the solution space, and they were also able to rank the inputs according to their influence on the GWP. Yet, the SA methods used in these studies have some shortcomings. Some of them only provide a qualitative ranking and fail to quantify the influence of each parameter on the GWP, while others cannot be applied on any type of model. Thus, a contradiction arises: How to find the right balance between increasing the sample size to explore a large solution space, and keeping a moderate sample size in order to avoid too expensive computational costs? This article suggests to couple SA and metamodels to overcome this issue. Indeed, after training on a reference dataset, a metamodel could be able to predict the GWP of new design alternatives, with a low computational time. It would then be possible to first generate a dataset with the minimum size for the SA and the metamodel training, and second to keep exploring a larger solution space thanks to the metamodel. The remaining questions are: Which SA and metamodel techniques to choose, and how much data does it need to predict the GWP accurately?

This article attempts to find answers to these questions. After reviewing several SA methods, one was selected, namely Sobol. It has been tested in order to determine if it is able to overcome the limits encountered with other methods used in previous studies. Then, three metamodels, namely Multiple Linear Regression (MLR), Support Vector Regression (SVR) and Artificial Neural Network (ANN), were chosen among several reviewed techniques. They are tested, with the aim of finding out if they are able to predict the GWP accurately, while being less computationally expensive than the original model. In addition to comparing different methods, this article also aims at assessing the initial sample size of design alternatives that is needed to perform accurate SA but also to train a robust metamodel. One can keep in mind that generating these data is computationally expensive, therefore the chosen sample size should be as small as possible. That is why the previously mentioned methods are tested on several datasets with different sizes. In order to increase the scalability of this study, the tests are conducted on a case study with a residential building and an office building, and the GWP characterizes the impacts of each building.

2. State-of-the-art

2.1. Sensitivity analysis in the literature

Sensitivity analysis is a statistical method that studies how the variation of inputs influences the value of the output. It can be used to rank inputs according to their influence on the output value. SA can be done by using different approaches, therefore several methods have been developed. In 2008, Saltelli wrote a book that summarizes the theoretical bases of SA as well as the different existing methods (Saltelli, 2008). In 2013, Tian reviewed some SA methods and described their respective advantages and drawbacks (Tian, 2013). More recently, Pianosi et al. proposed a review of the different methods, as well as some guidelines for users. They made a classification of SA methods based on the number of model evaluations, the purpose of the analysis, and the sampling approach (Pianosi et al., 2016). SA techniques can be divided in two categories: local methods and global methods.

The local approach (Saltelli, 2008) estimates derivatives at a specific point of the input space. It is the simplest type of SA method, and has a very low computational time. However, it only explores a reduced part of the input space (approximately D samples, where D is the number of inputs), as it only considers small variations of inputs around a specific point (Saltelli & Annoni, 2010). It is also only suitable for linear models and is not able to consider interactions between inputs.

In opposition to local methods, global methods consider the whole variation range of inputs. Global methods include screening methods, such as the Morris approach, regression-based methods, as well as variance-based methods like Sobol and FAST.

The Morris method, which was introduced by Morris in 1991 (Morris, 1991), enables to classify the inputs in three categories: those who have a negligible influence on the output, those who have a linear effect, and those who have non-linear and/or interaction effects. It is based on a One-At-A-Time sampling with random initializations of the different starting points. The Morris approach is a model-independent method, which means that no prior assumption is made about the model, and it has a low computational time. However, it only provides qualitative results, and allows a limited exploration of the input space (about 10 x D samples).

Regression-based methods measure sensitivity indices as coefficients of a linear regression. They include several methods such as Standardized Regression Coefficient (SRC) and Standardized Rank Regression Coefficients (SRRC). Correlation-based methods are similar but the sensitivity indices are computed with correlations between input and output. There are several correlation based methods such as Partial Correlation Coefficient (PCC) and Partial Rank Correlation Coefficient (PRCC) (Saltelli & Marivoet, 1990). Both regression and correlation based methods are easy to understand, have a low computational time and provide quantitative results. However, they can only be used with certain assumptions about the model. Indeed, SRC and PCC are only suitable for linear models, and SRRC and PRCC are suitable for non-linear models but not for non-monotonic models. Moreover, they are unable to take interactions between inputs into account. The exploration of the solution space is better than with the Morris method but remains limited (about 100 x D samples).

Variance-based methods provide indices computed thanks to a decomposition of the output's variance. First-order indices represent the influence of each variable individually, while total-order indices take interactions between variables into account. Variance-based methods have a higher computational time than the previous cited methods. However, they are model-independent and suitable for non-linear as well as non-monotonic models. They are also able to provide quantitative results and to give information about interactions between inputs. They need a higher number of simulations to perform the analysis and therefore the solution space is larger than the one generated with other methods (about 1000 x D samples). Variance-based methods include the methods of Sobol and FAST. The latter was introduced by Cukier, Fortuin, Shuler, Petschek, and Schaibly (1973). At first, it could only calculate first-order indices, but it was then improved into an « extended FAST » by Saltelli, Tarantola, and Chan (1999). This new version of FAST enables to calculate total-order indices but according to Tian (2013) FAST is not suitable for discrete distributions. The Sobol method was developed by the mathematician Sobol in 1993 (Sobol, 1993) and improved by Saltelli (2002). This method is able to compute both first-order indices and total-order indices. It also provides confidence intervals for the estimated values.

SA is widely used in the field of building performance. Indeed, it enables to identify the inputs with the highest impact on the output and therefore makes it easier for designers to focus on the most important variables. SA is also useful to reduce the number of dimensions of complex models by setting non-influential variables at constant values. Several studies about building energy performance are based on SA methods, they were reviewed by Tian (2013). However, SA was almost never applied on the GWP of buildings. We can nevertheless cite the work of three authors who performed a SA on the GWP of buildings. In 2013, Basbagill et al. (2013) proposed to integrate Building Information Modeling (BIM) software with new functionalities such as LCA and SA, with the aim of helping designers to understand which parameters are important for the GWP at the early design stage. They performed a SA