Sensitivity analysis and calibration of a numerical code for the prediction of power from a photovoltaic plant

Mathieu Carmassi *[1,2], Pierre Barbillon[1], Hassan Bouia[2], Merlin Keller[3], Amy Lindsay[2] , Éric Parent[1]

[1]UMR MIA-Paris, AgroParisTech, INRA, Université Paris-Saclay, Paris, France
[2]EDF R&D EnerBat department, Moret-sur-Loing, France
[3]EDF R&D MRI department, Chatou, France

April 22, 2016

*mathieu.carmassi@edf.fr
mathieu.carmassi@agroparistech.fr

Abstract

So far, most studies in the photovoltaic (PV) field have been done with a pseudo-deterministic point of view. The input uncertainty is propagated through a numerical code and provides the results with a 90% interval confidence, the parameters and the input probability distributions being determined by an expert. However, for investors, the risk associated to the investment is closely linked to the uncertainties in the evaluation of how much the PV power plant will produce. A statistical framework is needed to provide a more accurate estimation of the power produced by the PV plant and the associated error. For the modeling of PV power plants, a variety of computer models have been constructed. One of them, which is also the most accurate to date, is highly time-consuming. Hence, this study will deal with the sensitivity analysis and calibration of time-consuming codes, based on a large amount of data.

The modeling of a PV power plant can be performed with an equivalent electric scheme which tries to match its physical behavior. Two physical models are introduced. The first one is "simple" and the physical equations are straightforward. A Python code can be made up from those equations and predicts the amount of power, generated by a number of panels in simplified environmental conditions, relatively quickly. The second one is more complete. It matches better the real behavior of the PV power plant and can account for the partial shadings that may occur in a large-size power plant. Thus, the second physical model is more interesting to work with but also more complicated (high computing time and less regular behavior for example).

A physical model has two kinds of inputs: controlled variables which are observed in experimental conditions on the one hand and (usually uncertain) parameters on the other hand. The latter have to be calibrated to make the outputs of the physical model close to the observed quantities of interest, for instance the instantaneous power of the PV plant. Controlled variables are, for example, the meteorological data, consisting of: the amount of irradiation from the sun, the temperature, the geographical position (latitude and longitude) and the time. The parameters are factors inherent to the physical model (the yield of the PV module, the module temperature coefficient, etc.). Generally, the values of these factors are fixed according to expert elicitation. However, real-life experience shows that, when parameters are set according to expert opinion only, code outputs may be far from experimental data. To determine and evaluate the uncertainties on these parameters more precisely, a calibration has to be conducted. A Bayesian framework is adopted to make this inference. This will allow us to confront expert information and experimental data.

The first step for such a study is to perform a sensitivity analysis. This is crucial for the following step because sorting parameters from the most to the less important will allow us to save a lot of computational time thereafter. The number of parameters to be calibrated depends on the physical model. However in both present cases it always exceeds ten. A screening method is first carried out to separate the ones which have no overall impact on the output. Afterwards, a Sobol analysis is done to sort the remaining parameters and indicate which one is the most important. The Sobol analysis is important in this case, because it provides a physical point of

view on which parameters have an impact on the power and allows us to check the accuracy of the analysis.

The second step is to calibrate the parameters. Once a statistical model is set, Bayesian inference will combine all available data with a *prior* distribution to obtain a *posterior* distribution on the unknown parameters. The experimental data are the power measured instantaneously on the test stand. At a high sampling frequency, the number of experimental points is very high. However, all of them are not really informative. For example, during the night the PV power plant produces nothing and these data can be removed from the data set. Globally, the considered data can be limited to the time period from daybreak to sunset. The calibration is performed with Markov Chains Monte Carlo algorithms (MCMC) like Metropolis-Hastings or Metropolis within Gibbs algorithms. A tempering scheme is adopted to deal with the full amount of available data without jeopardizing the time needed to achieve convergence of MCMC algorithms. Furthermore, an adaptive exploration kernel is chosen for the MCMC algorithms since the dimension is high and adapting the exploration kernel to the covariance of the posterior distribution will accelerate convergence.

To ensure that the production predictions of the power plant provided by the code are reliable, the code has to be validated. Validation means to assess whether the code produces outputs close to observed power measures once calibration has been conducted. This validation question can be expressed as a choice between two statistical models. In the first one, the only error between the outputs of the physical model and the observed power measures is a measurement error i.e. a classical white noise process. In the second one, a discrepancy term will be added to the measurement error to capture a systematic error of the physical model. Usually, this discrepancy is modeled as a Gaussian process. If the model with the discrepancy is selected by the statistical procedure, it will mean that the physical model does not predict well enough the actual power. In this case, pure model predictions can be corrected by adding the estimated discrepancy.

Finally, all of these techniques are computationally demanding. Indeed, some of the physical models are expensive and a Gaussian process emulator has been introduced to reduce the overall computation load.

Keywords: Sensitivity analysis, Photovoltaic power plant, Morris screening method, Sobol indices, tempering scheme, Computer code, Bayesian calibration.