An Application of Calibration and Uncertainty Quantification Techniques for Agent-Based Models

Sotiris Papadelis and Alexandros Flamos

Abstract In this chapter, a step-by-step application of calibrating an agent-based model is presented. In particular, an agent-based model for small-scale PV adoption was calibrated on the historical data for the small-scale solar PV capacity additions that took place in Greece from January 2010 to February 2013. The process of the model calibration allowed to (a) quantify and take into consideration uncertainties that are related to the characteristics and the decision-making criteria of the agents (i.e. independent PV power producers), in contrast to the more obvious uncertainties, such as technology costs, and (b) use the calibration results to explore the plausible—given the historical data—behaviour of the potential PV adopters in Greece under the new net-metering scheme (in effect as of mid-2015).

Keywords Agent-based · Calibration · Stochastic emulators · History matching · Uncertainty quantification

1 Introduction

Models try to narrow the differences between decision-makers' thinking, reasoning, representation and computing (Doukas 2013). Anytime we decide to develop a model, we have two choices. The first one is to rely on physical laws that we know they hold true or mathematical relationships that are valid as long as the underlying theory is valid too. The second choice is to come up with a plausible model that captures relationships and dynamics that we expect to be true and acknowledge the fact that many parts of the model will inevitably be arbitrary.

Agent-based models (ABMs) fall into the second choice (Flamos 2016). This means that developing a new ABM is an interesting endeavour, but unless we

S. Papadelis · A. Flamos (✉)

Technoeconomics of Energy Systems Laboratory (TEESLab), Department of Industrial Management & Technology, University of Piraeus (UNIPi), Piraeus, Greece
e-mail: aflamos@unipi.gr

identify the uncertainty that governs the applicability of the model, it is just a modelling exercise. While ABMs give us increased flexibility in modelling complex systems driven by agent actions and interactions, they give rise to the question of when an ABM represents reality well enough to extrapolate into the future.

A common approach to validating the extent at which an ABM is realistic is through calibration using historical data. If done right, this process is an opportunity to deal with model uncertainty, which is the specific type of uncertainty that stems from the fact that there exist many variations of an ABM that are all plausible under the same set of historical data.

Agreement between model and data does not imply that the modelling assumptions accurately describe the processes producing the observed behaviour; it merely indicates that the model is one (of maybe several) that is plausible. As a result, dealing with model uncertainty during calibration allows us to address the trade-off between a very flexible model that can easily fit on the historical data and model uncertainty.

In this chapter, we present a method for quantifying model uncertainty. The ABM used aims at replicating the dynamics of small-scale PV adoption in Greece. Its outputs are determined by the values of its ten free parameters. As a result, before the model can be used for forward-looking simulations, these parameters must be constrained into a subspace of plausible values.

Accordingly, the model was calibrated on the data for the period from 2010 to 2013. This period was the period with the largest additions in small-scale PV capacity (Anagnostopoulos et al. 2017; Papadelis et al. 2016). The calibration data corresponds to:

- a) Past prices and market conditions (prices for small-scale PV systems and the relevant Greek feed-in tariff (FIT) prices)
- b) The requests for grid connections, the records for which are available at the individual request level¹

The chapter is structured as follows: Sect. 2 presents a brief description of the employed ABM, Sect. 3 presents the concept of Gaussian process emulators that play a central role in the calibration method that was applied, Sect. 4 presents the details of fitting a Gaussian process emulator on the ABM results and Sect. 5 presents the method and results of the model calibration. The chapter concludes with a discussion.

2 The ABM for the Diffusion of Small-Scale Solar PV

The parameters that govern the model's behaviour are as follows:

1. **Agent initial beliefs.** Each agent has a private initial belief for the expected annual cash inflows from investing in 300 Wp of solar PV. This belief is

¹http://www.rae.gr/site/categories_new/regirsty/licences.csp

expressed as a Gaussian distribution with a mean value μ^{CF} and a precision ρ^{CF} . Low values of ρ^{CF} reflect flexibility in beliefs (i.e. little evidence is enough to shift the agent's beliefs). Low values of μ^{CF} reflect pessimism about the payoffs of the investment. The initial beliefs of each agent are randomly drawn from two global probability distributions—one for the mean value μ^{CF} and one for the precision ρ^{CF} of each agent—while the parameters of both distributions are regarded as parameters that need to be set through calibration.

2. **Social learning.** To capture the effects of social learning, each agent receives information from the agents in its social circle that have already invested in PV. This information concerns the actual profitability of their investments so far, and it is used to update² the agent's beliefs. Accordingly, each agent has a social circle. This condition is modelled as a “small-world” network (Watts and Strogatz 1998).
3. **Resistance towards PV investments.** Each agent is characterized by their resistance towards investing in solar PV. Resistance is defined as a weighted sum of two parameters:

- (a) The profitability of the investment expressed by its payback period, so that the larger the profitability, the shorter the payback period, and, thus, the lower the resistance.

We assume that agents are able to use their beliefs regarding the expected cash inflows to estimate the profitability of investing. Given the fact that the agents' beliefs are expressed probabilistically, we can calculate the following z_i value, where $i = 1, 2, \dots$ represents the years from the simulation's “now”:

$$z_i = \frac{\mu^{\text{CF}} \cdot \sum_{t=1}^i \left(\frac{1}{(1+d)^t} \right) - \text{capex}}{i/\rho^{\text{CF}}} \sim N(0, 1)$$

where **capex** is the capital expenditure and d the discount rate.

The term z_i gives us the probability that the (discounted) payback period of the investment is equal to i . Making the assumption that all agents evaluate an investment based on when it will have paid itself back with probability 90%; this formula gives us the respective payback period i .

- (b) The difference between the total number of agents in the simulation and the number of them that has already invested in PV—so that the larger the installed base, the smaller the resistance.

Baranzini et al. (2017) provide evidence that both learning and imitation are important components of social contagion for the adoption of solar PV technology. By making attitude towards PV a function of the installed base, we aim to capture the imitation (social influence) aspect.

²This is equivalent to updating a Gaussian prior given a new observation.

Fig. 1 Size distribution of grid connection requests for small-scale PV systems in Greece

The weights of this sum are derived from a global probability distribution, the parameters of which are regarded as parameters that need to be set through calibration.

4. **Propensity to invest.** Each agent has a threshold value for their resistance parameter. When the latter falls under the threshold value, action could—but not necessarily will—be induced. For the calibration phase, we assumed that when agents decide to invest in a PV system, its size is given by the empirical probability distribution that was derived from the available historical data (Fig. 1).

Finally, Young (2009) notes that inertia is the simplest reason why innovations take time to diffuse: people delay in acting on new information. Accordingly, we have included inertia in the model by defining a global parameter (i.e. it has the same value for all agents in the model) that represents the probability that any agent would actually invest if their resistance threshold is crossed.

3 The Concept of Emulators

An emulator is a probabilistic approximation of the ABM. The probabilistic nature of the emulators makes them ideal for the quantification of the uncertainty regarding their estimations, as well as of the way the uncertainty of the ABM parameters gets propagated into its results. In addition, emulators can be evaluated significantly faster in comparison to the actual model, which allows us to employ Monte Carlo sampling methods that would be otherwise prohibitively expensive in terms of computation resources.

In the application presented in this chapter, Gaussian processes are used as emulators. Accordingly, a brief description of Gaussian processes is provided.

3.1 Gaussian Processes for Regression

Suppose that we have some function $f(x)$ and we want to estimate the value of this function at the input points x_1, x_2 and x_3 . As long as we believe that a suitable prior of each variable $f(x)$ could follow a Gaussian distribution, the marginalization property of the Gaussian allows us to write

$$\mathbf{f} = \begin{bmatrix} f(x_1) \\ f(x_2) \\ f(x_3) \end{bmatrix} \sim N \left(\begin{bmatrix} f(x_1) \\ f(x_2) \\ f(x_3) \end{bmatrix} \mid \begin{bmatrix} m_f(x_1) \\ m_f(x_2) \\ m_f(x_3) \end{bmatrix}, \begin{bmatrix} \sigma_f^2(x_1) & 0 & 0 \\ 0 & \sigma_f^2(x_2) & 0 \\ 0 & 0 & \sigma_f^2(x_3) \end{bmatrix} \right) \quad (1)$$

where $m_f(x_i)$ is the mean value of $f(x_i)$ and $\sigma_f^2(x_i)$ its variance.

If we also believe that the values of this function are related to each other (i.e. knowing one can tell us something about the others), we can assume that the function values follow a joint Gaussian distribution that imposes covariance between these values. Covariance is introduced by defining a covariance function $k_f(\mathbf{x}, \mathbf{x}^*)$ that models the covariance between the function values at different inputs $\mathbf{x}$ and $\mathbf{x}^*$:

$$k_f(\mathbf{x}, \mathbf{x}^*) = \text{Cov}[f(\mathbf{x}), f(\mathbf{x}^*)] \quad (2)$$

For the case of the three-dimensional Gaussian distribution of this example, the prior distribution becomes now of the form:

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} f(x_1) \\ f(x_2) \\ f(x_3) \end{bmatrix} \sim N \left(\begin{bmatrix} f(x_1) \\ f(x_2) \\ f(x_3) \end{bmatrix} \mid \begin{bmatrix} m_f(x_1) \\ m_f(x_2) \\ m_f(x_3) \end{bmatrix}, \begin{bmatrix} k_f(x_1, x_1) & k_f(x_1, x_2) & k_f(x_1, x_3) \\ k_f(x_2, x_1) & k_f(x_2, x_2) & k_f(x_2, x_3) \\ k_f(x_3, x_1) & k_f(x_3, x_2) & k_f(x_3, x_3) \end{bmatrix} \right) \quad (3)$$

If we write $\mathbf{x} = [x_1, x_2, x_3]^T$, the distribution in Eq. (3) can be written in a more compact form as:

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} f(x_1) \\ f(x_2) \\ f(x_3) \end{bmatrix} \sim N(\mathbf{f} \mid m_f(\mathbf{x}), K_f(\mathbf{x}, \mathbf{x})) \quad (4)$$

where the mean function $m_f(\mathbf{x})$ reflects the expected function value at input $\mathbf{x}$:

$$m_f(\mathbf{x}) = E[f(\mathbf{x})] \quad (5)$$

A *Gaussian process* (GP) generalizes the aforementioned multivariate normal to infinite dimensions. It is defined as a collection of random variables f , any finite number of which has a joint Gaussian distribution (Rasmussen and Williams 2006). If we assume a GP prior over a set of function values f , we actually assume that:

1. **Their marginal distribution is Gaussian.** For any $\mathbf{x} = \mathbf{x}_i$, the prior distribution for $f(\mathbf{x}_i)$ is

$$f(\mathbf{x}_i) \sim N(m_f(\mathbf{x}_i), k_f(\mathbf{x}_i, \mathbf{x}_i)) \quad (6)$$

The covariance function k_f is called the kernel of the GP.

2. **Their joint distribution is also Gaussian.** For a sample $\mathbf{X} = \{\mathbf{x}_i, i = 1, \dots, n\}$ consisting of n pairs of inputs, the probability density function of the f 's values is

$$p(f|\mathbf{X}) = N(f | m_f(\mathbf{X}), K_{ff}(\mathbf{X})) \quad (7)$$

where $K_{ff}(\mathbf{X})$ is the covariance matrix obtained by evaluating the kernel k_f on all data points in $\mathbf{X}$:

$$K_{ff}(\mathbf{X}) = \begin{bmatrix} k_f(\mathbf{x}_1, \mathbf{x}_1) & k_f(\mathbf{x}_1, \mathbf{x}_2) & \cdots & k_f(\mathbf{x}_1, \mathbf{x}_n) \\ k_f(\mathbf{x}_2, \mathbf{x}_1) & k_f(\mathbf{x}_2, \mathbf{x}_2) & \cdots & k_f(\mathbf{x}_2, \mathbf{x}_n) \\ \vdots & \vdots & \ddots & \vdots \\ k_f(\mathbf{x}_n, \mathbf{x}_1) & k_f(\mathbf{x}_n, \mathbf{x}_2) & \cdots & k_f(\mathbf{x}_n, \mathbf{x}_n) \end{bmatrix} \quad (8)$$

Once a mean function and a kernel have been chosen, we have actually decided on a prior for the function f . This is a distribution over possible functions. Since it is not conditioned on any observed data, it represents our prior assumption of the function space from which the data may have been generated.

When we do observe some actual data $\mathbf{D} = \{(\mathbf{x}_{i,:}, y_i), i = 1, \dots, n\}$, the GP prior is conditioned on $\mathbf{D}$ to obtain a posterior GP process that represents all functions that are consistent with both the observed data and the prior. This posterior GP process $p(f|\mathbf{D})$ is used as a proxy for the original ABM.

The main advantage from using a GP is that it becomes analytically tractable to estimate the values of the approximated function on inputs that are not part of the observed data. In particular, we can make predictions for new inputs $\mathbf{X}_*$ by drawing f_* from the posterior $p(f|\mathbf{D})$. This posterior predictive distribution can be written as³:

$$p(f_*|\mathbf{D}) = N\left(f_* | m_f(\mathbf{X}_*) + K_{*f}K_{ff}^{-1}(y - m_f(\mathbf{X})), K_{**} - K_{*f}K_{ff}^{-1}K_{f*}\right) \quad (9)$$

³ Assuming that there are no observation errors.

This means that we can fit a GP on the results from a small number of parameter combinations for the ABM and then use the GP's posterior predictive distribution to estimate the model's results for other parameter combinations without a need to re-run the ABM.

3.2 *Benefits of Using Gaussian Processes as Emulators*

An important implication is that we have a formula that calculates the predictive uncertainty at each test input X_* :

$$\text{Var}(f|X_*) = K_{**} - K_{*f}K_{ff}^{-1}K_{f*} \quad (10)$$

This allows us to employ a sequential sampling scheme:

1. The GP emulator is fitted on the results from a small number of parameter combinations at which the ABM is run.
2. Additional parameter combinations for the ABM can be chosen corresponding to regions where the emulator's predictive uncertainty is high.
3. The GP emulator is updated to include the new ABM results.

The function $\text{Var}(f|X_*)$ can accept matrices as inputs. This allows us to evaluate it on a large set of possible inputs and then either select only the inputs (i.e. matrix rows) that correspond to a predictive uncertainty that is higher than a predetermined threshold or select a predetermined number of inputs that correspond to the highest predictive uncertainty.

4 The Design and Validation of the GP Emulator

4.1 *Options for the Emulator Form*

Following Haylock and O'Hagan (1996), the general form for the emulator to fit on the model output j is

$$g_j(\mathbf{x}) = \sum_{i=1}^q h_i(\mathbf{x})\beta_i + \mathbf{u}(\mathbf{x}) \quad (11)$$

The first term is a regression term, where $h_i(\cdot)$ are deterministic functions of the model parameters and β_i the regression coefficients. The second term is a GP with a zero mean function and an appropriately chosen kernel. One way to interpret this formulation is that the ABM is approximated by a regression term, which is then

improved by a GP. Alternatively, this formulation can be interpreted as an emulator that follows a GP that is given by:

$$p(g_j(\mathbf{x})) = N\left(\mathbf{x} \mid \sum_{i=1}^q h_i(\mathbf{x})\beta_i, k_f(\mathbf{x}, \mathbf{x} \mid \boldsymbol{\theta})\right) \quad (12)$$

where $\boldsymbol{\theta}$ denotes the parameters of the kernel.

The most common approach for estimating the parameters of the GP emulator (i.e. the coefficients β_i and the parameters $\boldsymbol{\theta}$ of the kernel) is the “plug-in” approach of Kennedy and O’Hagan (2001), where the parameters are estimated using maximum likelihood and, then, treated as fixed during the calibration of the model.

As far as the GP kernels are concerned, the most popular choice is the squared exponential kernel, which implies that the function to approximate has infinitely many derivatives. If the smoothness assumption of the squared exponential kernel is unrealistic for a specific modelling case, the Matérn kernel class can be used as an alternative (Genton 2002).

Both the squared exponential kernel and the kernels of the Matérn class include the following term:

$$k(\mathbf{x}_i, \mathbf{x}_j) \propto \frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{\ell} \quad (13)$$

The length scale ℓ determines how close two data points x_i, x_j have to be to influence each other significantly. If we allow for different values of ℓ per input dimension, then we achieve automatic relevance determination, so named because estimating the length scale parameters $\ell_1, \ell_2, \dots, \ell_q$ implicitly determines the “relevance” of each dimension. Input dimensions with relatively large length scales imply relatively little variation/impact on the function being approximated.

4.2 Fitting the GP Emulator

In the application presented in this chapter, a GP emulator—the STatistical approximation-based model EMulator (STEEM) of the Technoeconomics of Energy Systems Laboratory (TEESlab)—was fitted on the results from a number of parameter combinations at which the ABM was run. This means that the parameter combinations were the input data and the ABM results were the output data for the GP emulator. The mean function of the emulator was set to zero, while the Matérn 3/2 kernel was used as the covariance function.

To generate the input data, preliminary ranges for the value of each model parameter were derived, and then the ABM was run for 150 different parameter combinations. These initial combinations were chosen using a maximin Latin

Table 1 ABM parameter description and initial ranges

	Description	Min	Max
1.	The shape parameter of the global distribution that assigns μ^{CF} to each agent in the model	100	250
2.	The scale parameter of the global distribution that assigns μ^{CF} to each agent in the model	10	50
3.	The shape parameter of the global distribution that assigns ρ^{CF} to each agent in the model	10	50
4.	The scale parameter of the global distribution that assigns ρ^{CF} to each agent in the model	5	20
5.	The shape parameter of the global distribution that assigns the weight of the profitability to each agent's resistance	0.5	5
6.	The scale parameter of the global distribution that assigns the weight of the profitability to each agent's resistance	0.1	1
7.	The shape parameter of the global distribution that assigns the weight of the installed base to each agent's resistance	0.5	5
8.	The scale parameter of the global distribution that assigns the weight of the installed base to each agent's resistance	0.1	1
9.	The shape parameter of the global distribution that assigns each agent's threshold value for their resistance parameter	10	30
10.	The scale parameter of the global distribution that assigns each agent's threshold value for their resistance parameter	5	10

hypercube design, to fill the entire input space by maximizing the minimum distance between the points generated.

The initial ranges for the ABM parameters were defined as presented in Table 1. There are two possible caveats when selecting the initial ranges for the model parameters to be calibrated:

1. The ranges are wider than necessary. Sensitivity analysis will be carried out to test if and for which parameter this is true.
2. The ranges are too narrow. At the end of the calibration, one should revisit these ranges.

The corresponding PV capacity pathways were scaled to the [0, 1] range. This allowed the comparison of the simulated pathways with the—also scaled—pathway of the actual PV investments during the calibration period.

The ABM calibration was conducted with the assumption that the model represents reality well if it can replicate the historical growth rates for PV installations. In other words, calibration looked for similar shapes of the cumulative capacity of PV installations, while the scale is inevitably different. At the same time, calibrating based only on the shape of the simulated pathways creates a problem of non-identifiability; we need a final target to use as a point of reference. To this end, the median of the capacities achieved at the end of the simulated period was mapped to the value of 1 during the scaling. In addition, the input data was normalized.

The ABM results were split into four periods (mid-2010, end-2010, end-2011, end of simulation period), and four GP emulators (four implementations of the STEEM) were fitted on the respective results. The hyperparameters of each emulator were estimated using likelihood maximization, and the Python package GPy⁴ was employed for this purpose.

4.3 Diagnostics

Bastos and O'Hagan (2009) present different validation methods for GP emulators. In this chapter, the generalization capability of the fitted emulator was judged by performing k-fold cross-validation on the emulator that corresponds to the results at the end of the simulated period. K-fold cross-validation was selected to make sure that there is no region in the input space that the emulator has been poorly fitted. As a metric for the validation, the coefficient of Nash-Sutcliffe efficiency was used, which is given by:

$$E = 1 - \frac{\sum_{i=1}^m (y_i - y_{i,GP})^2}{\sum_{i=1}^m (y_i - \bar{y})^2} \quad (14)$$

where m represents the number of data in each fold, y_i are the respective outputs of the ABM, $\bar{y}$ is the mean of these outputs and $y_{i,GP}$ are the estimations of the GP emulator for the same parameter inputs. This coefficient has an optimal value of 1; a value of 0 indicates that a forecast using the mean of the ABM outputs will have the same utility as the result of the GP emulator.

4.4 Sensitivity Analysis

The preliminary ranges for the value of each model parameter were chosen arbitrarily with the goal of fitting a GP emulator on a reasonably large input space. However, after the emulator is fitted, it is useful to revisit the ranges by performing a sensitivity analysis that is facilitated by the emulator. Sensitivity analysis explains which of the d features in the input dataset $\mathbf{X} = \{\mathbf{x}_j, j = 1, \dots, d\}$ are most responsible for the uncertainty in the model's result. It makes sense to restrict the range of the parameters that have only a small impact on the uncertainty of the model's outputs.

For each feature $j : 1, 2, \dots, d$, we split the dataset $\mathbf{X}$ into two parts, one including the selected feature and the other the remaining ones:

⁴<https://github.com/SheffieldML/GPy/>

$$E\mathbf{X} = (\mathbf{x}_j, \mathbf{X}_{-j}) \quad (15)$$

Then, we can assess the sensitivity of the model's output to the uncertainty regarding $\mathbf{x}_j$ through the expected reduction in the output's variance if the true value of $\mathbf{x}_j$ is learnt. Saltelli et al. (2010) provide the following formula for the calculation of the expected reduction in variance:

$$\text{Var}_{\mathbf{x}_j}(E(y | \mathbf{x}_j)) = \frac{1}{N} \sum_{i=1}^N \left\{ f(\mathbf{B})_i \left(f(\mathbf{A}_B^{(j)})_i - f(\mathbf{A})_i \right) \right\} \quad (16)$$

$\mathbf{A}$ and $\mathbf{B}$ are two independent $N \times d$ matrices that contain samples of the model's inputs. Following Saltelli et al. (2010), a Sobol sequence was used to derive these matrices. Sobol sequences were designed to cover the unit hypercube with lower discrepancy than completely random sampling. The index j runs from 1 to d , where d is the number of parameters. The index i runs from 1 to N , where N is the number of input samples. The term $f(\mathbf{A})_i$ represents the i th element of the vector that is the output of the GP emulator when evaluated at $\mathbf{X}_* = \mathbf{A}$. The term $\mathbf{A}_B^{(j)}$ represents a matrix where column j comes from matrix $\mathbf{B}$ and all other columns come from matrix $\mathbf{A}$. The matrices $\mathbf{A}$ and $\mathbf{B}$ can be generated from a Sobol sequence of size $N \times 2d$: $\mathbf{A}$ is the left half of the sequence and $\mathbf{B}$ is the right part of it.

Given $\text{Var}_{\mathbf{x}_j}(E(y | \mathbf{x}_j))$, we can compute the first-order sensitivity coefficient S_j that captures the main effect of x_j on y as:

$$S_j = \frac{\text{Var}_{\mathbf{x}_j}(E(y | \mathbf{x}_j))}{\text{Var}(y)} \in [0, 1] \quad (17)$$

As it turns out, the ABM's output variance is mainly driven by:

1. The shape parameter of the global distribution that assigns each agent's threshold value for their resistance parameter
2. The shape parameter of the global distribution that assigns the weight of the profitability to each agent's resistance
3. The shape parameter of the global distribution that assigns μ^{CF} to each agent in the model

Accordingly, all other parameters were fixed to the mean value of their initial ranges.

5 Model Calibration

5.1 The History Matching Method

The calibration of the model was based on the history matching method (Craig et al. 1997). History matching works by excluding subsets of the model's parameter

space, which are unlikely to provide a good match between the model outputs and the observed reality. As the plausible space becomes smaller and smaller, emulators become smoother and more accurate, allowing us to zoom in the parameter space we explore. As a result, the implausible parameter values are excluded in iterations, known as waves, where new emulators are built after each wave. History matching has been successfully applied across a wide range of scientific domains, including the case of calibrating ABM (Andrianakis et al. 2015).

A central concept of the history matching method is the quantification of the major uncertainties that have an impact on the calibration process (Kennedy and O'Hagan 2001):

1. **Ensemble variability** (V_{es}). This uncertainty represents the stochastic nature of the ABM; different runs with the same parameters will give different outcomes. To calculate the ensemble variability, one combination for the model's parameters was selected—it was assumed that the variability is independent from the inputs, so one combination will suffice—and run the model $K = 25$ times to capture the variability of the results. Accordingly, the ensemble variability was calculated to be $V_{es}=0.002$.
2. **Emulation uncertainty** (V_{em}). This type of uncertainty originates in the possible discrepancy between the ABM and the GP-based STEEM. To calculate the emulation uncertainty, the model runs used for the calculation of the ensemble uncertainty were utilized. The emulation uncertainty was given by:

$$V_{em} = \frac{1}{K-1} \sum_{k=1}^K (f_{j,k}(x) - g_j(x))^2 \quad (18)$$

where $f_{j,k}(x)$ is the output of the ABM for the period j at the k th run and $g_j(x)$ is the emulator's result. Accordingly, the emulation uncertainty was calculated to be $V_{em} = 0.002$.

Having specified the aforementioned uncertainties, the implausibility of a specific parameter combination $\mathbf{x}$ is given by (Vernon et al. 2010):

$$I_j(\mathbf{x}) = \frac{|z_j - g_j(\mathbf{x})|}{\sqrt{V_{es} + V_{em}}} \quad (19)$$

where z_j is the observed data for the period j .

A natural cut-off value for implausibility is 3, i.e. any parameter combination $\mathbf{x}$ with $I(\mathbf{x}) > 3$ should be considered implausible.

For the calculation of the implausibility function, the following steps took place:

1. $z_j \rightarrow$ The actual data was scaled to the $[0, 1]$ range.
2. $\mathbf{x} \rightarrow$ A new and much larger (20,000) set of combinations was derived through Latin hypercube design, this time taking random points within the sampling intervals.

3. $g_j(\mathbf{x}) \rightarrow$ We predicted the cumulative capacity additions at the end of the period j for each combination.
4. $I_j(\mathbf{x}) \rightarrow$ The implausibility function was used as an indicator function to split the samples into plausible and implausible ones through the Patient Rule Induction Method (Friedman and Fisher 1999).

5.2 The Patient Rule Induction Method

The Patient Rule Induction Method (PRIM) is a heuristic algorithm that aims to identify rectangular partitions (called boxes) of the input space where the average response is much larger than the average response across the whole input space. When the responses correspond to classification classes, the goal is to identify partitions with high homogeneity or, equivalently, low impurity as measured by the Gini index:

$$GI = 1 - \sum_i p_i^2 \quad (20)$$

where p_i is the fraction of responses labelled with class i in the partition.

For binary responses (as is our case), both aforementioned goals are equivalent if the responses of no interest (i.e. with implausibility function values greater than 3) are assigned the zero value.

The PRIM algorithm proceeds as follows:

1. Identify a rectangle box that includes all input data, B_1 .

Identify a dimension j that when removing the slice of data below the a quantile or above the $1 - a$ quantile of the corresponding variable's distribution, we achieve the greatest decrease in the impurity of the remaining data. The parameter a is user-defined.

The objective function that we used divides the gain in purity by the loss in mass:

$$\max_j \frac{GI_{B_1} - GI_{B_1-j}}{\beta_{B_1} - \beta_{B_1-j}} \quad (21)$$

where GI_{B_1} is the Gini index of the initial box B_1 and GI_{B_1-j} the Gini index of the box resulting from removing the amount of β_{B_1-j} data from dimension j .

2. Repeat step 2 until a minimum number β of data remains in the rectangle. This process is called "peeling". The parameter β is also user-defined.
3. Reverse the peeling process by expanding the rectangle in the dimension where adding a slice of data leads to the greatest decrease in the impurity of the expanded partition.

4. Repeat step 4 until no decrease in impurity is possible. This process is called “pasting” and aims to refine the box boundaries.

Based on the aforementioned, PRIM was applied so as to find boxes that contain plausible parameter combinations (i.e. with $I(\mathbf{x}) \leq 3$). The quality of the derived partition was judged by measuring its density. Density is the ratio of the total number of plausible combinations inside the box to the number of all combinations inside this box.

In addition, the partition boxes were evaluated according to their coverage. Coverage is the ratio of the total number of plausible combinations inside the box to the total number of plausible combinations found across the whole dataset. If the coverage is low, it is possible that either the data has high levels of noise or there are additional partitions with high density as well.

To explore the latter possibility, the method of Guivarch et al. (2016) can be used, according to which the data in the identified box are marked as uninteresting (i.e. they are assigned the zero value), and the PRIM algorithm is re-run so as to identify a new box if possible.

5.3 Calibration and Extrapolation Results

After considering all periods, the PRIM constrained the parameters 1, 5 and 9 of Table 1 as presented in Table 2.

Subsequently, the calibrated model was used to explore the expected effectiveness of the Greek net-metering scheme in driving investments in small-scale PV during the period 2018–2025. To this end, a set of scenarios for different plausible values of the model parameters was run, assuming that the retail prices remained unaffected and taking into consideration the provisions of the Greek net-metering scheme.

The ABM that was used for the forward-looking simulations differs from the one used for calibration in the following ways:

1. **Available options.** When the agents decide to invest in solar PV, they can choose one option from a limited set of different ones, all of which are available to all

Table 2 Plausible ranges for the ABM parameters after calibration

	Description	After calibration	
		Min	Max
1.	The shape parameter of the global distribution that assigns μ^{CF} to each agent in the model	181	192
5.	The shape parameter of the global distribution that assigns the weight of the profitability to each agent’s resistance	1.9	3
9.	The shape parameter of the global distribution that assigns each agent’s threshold value for their resistance parameter	11.6	19.6

Fig. 2 New PV capacity additions due to the net-metering scheme

agents. These options concern the size of the PV system installed: 2.4 kWp, 4.8 kWp or 9.6 kWp.

2. **The option selection rule.** If more than one option is evaluated favourably by an agent, the selection is based on the softmax rule; the probability of investing in option j is related to the agent's resistance to it (r_j) according to:

$$P(j) = \frac{e^{-r_j}}{\sum_k e^{-r_k}}$$

The results are presented in Fig. 2.

6 Discussion

The ambition of this chapter was to highlight the fact that when calibrating ABMs to historical data, we should aim to explicitly quantify model uncertainty. For the model under study, this uncertainty can be regarded as uncertainty that is related to the characteristics and the decision-making criteria of the agents (i.e. independent PV power producers), and it is evident by the range of the different plausible parameters of the model and the range in the results that they produce.

The main takeaway is that model uncertainty exists whether we choose to quantify it or not. However, by quantifying it, we are able to understand the utility and the limits of our model. In this case, we can compare the future scenarios with the historical data (Fig. 3). This reveals that the expected results from the net-metering scheme in Greece are positive but not as strong as the results of the FIT period; even with the most favourable of the plausible parameter combinations, it will take 7 years to achieve the same PV capacity additions that the FIT scheme achieved from 2010 up to the end of 2012.

Fig. 3 Grid connection requests for small-scale PV systems in Greece

In addition, highlighting model uncertainty makes it sensible that a policy design process that utilizes ABMs is structured around the concept of adaptability; as new data on the actual decisions of the relevant actors is accumulated, the initial policy should adapt, the same way as it should adapt to changes in its environment, such as technology costs.

Finally, ABMs can help policies succeed by focusing on directly affecting the agents' characteristics. As an example, dissemination of trusted information can help align agents' beliefs, while marketing campaigns and appropriately devised narratives can help increase the degree of influence by social learning and imitation.

Acknowledgements This research was funded by the EC Horizon 2020 Project titled “Transitions pathways and risk analysis for climate change mitigation and adaption strategies” (TRANSrisk) with Grant Agreement No. 642260. The authors would like to acknowledge the support from the EC. The content of the paper is the sole responsibility of its authors and does not necessary reflect the views of the EC.

References

- Anagnostopoulos, P., Spyridaki, N.-A., & Flamos, A. (2017). A “New-Deal” for the development of photovoltaic investments in Greece? A parametric techno-economic assessment. *Energies*, 10 (8), 1173.
- Andrianakis, I., Vernon, I. R., McCreesh, N., McKinley, T. J., Oakley, J. E., et al. (2015). Bayesian history matching of complex infectious disease models using emulation: A tutorial and a case study on HIV in Uganda. *PLOS Computational Biology*, 11(1), e1003968.

- Baranzini, A., Carattini, S., & Peclat, M. (2017). *What drives social contagion in the adoption of solar photovoltaic technology*. GRI Working Papers 270, Grantham Research Institute on Climate Change and the Environment.
- Bastos, L. S., & O'Hagan, A. (2009). Diagnostics for Gaussian process emulators. *Technometrics*, 51, 425–438.
- Craig, P. S., Goldstein, M., Seheult, A. H., & Smith, J. A. (1997). Pressure matching for hydro-carbon reservoirs: A case study in the use of Bayes linear strategies for large computer experiments (with discussion). In C. Gastonis et al. (Eds.), *Case studies in Bayesian statistics* (Vol. III, pp. 37–93). New York: Springer.
- Doukas, H. (2013). Linguistic multicriteria decision-making for energy systems: Building the “RE₂S” framework. *Wiley Interdisciplinary Reviews Energy and Environment (WIREs)*, 2(5), 571–585.
- Flamos, A. (2016). A sectoral micro-economic approach to scenario selection and development: The case of the Greek power sector. *Energies*, 9, 77.
- Friedman, J. H., & Fisher, N. I. (1999). Bump hunting in high-dimensional data. *Statistics and Computing*, 9(2), 123–143.
- Genton, M. G. (2002). Classes of kernels for machine learning: A statistics perspective. *Journal of Machine Learning Research*, 2, 299–312.
- Guivarch, C., Rozenberg, J., & Schweizer, V. (2016). The diversity of socio-economic pathways and CO₂ emissions scenarios: Insights from the investigation of a scenarios database. *Environmental Modelling & Software*, 80, 336–353.
- Haylock, R. G., & O'Hagan, A. (1996). On inference for outputs of computationally expensive algorithms with uncertainty on the inputs. *Bayesian Statistics*, 5, 629–637.
- Kennedy, M. C., & O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63, 425–464.
- Papadelis, S., Stavarakas, V., & Flamos, A. (2016). What do capacity deployment rates tell us about the efficiency of RES-E support measures in Greece? *Energies*, 9(1), 38.
- Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning*. Cambridge, MA: MIT Press.
- Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., & Tarantola, S. (2010). Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2), 259–270.
- Vernon, I., Goldstein, M., & Bower, R. G. (2010). Galaxy formation: A Bayesian uncertainty analysis. *Bayesian Analysis*, 5(4), 619–670.
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of ‘small-world’ networks. *Nature*, 393, 440–442.
- Young, H. (2009). Innovation diffusion in heterogeneous populations: Contagion, social influence, and social learning. *American Economic Review*, 99(5), 1899–1924.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.