

A Baseline Calibration Procedure for CGE models: An Application for DART

Ding Jin^a, Sneha Thube^b, Johannes Hedtrich^b, Christian Henning^b
and Ruth Delzeit^b

¹Kiel Institute for the World Economy, Germany

²Institute of Agricultural Economics, University of Kiel, Germany

Abstract

In the recent years the research interests in the field of Computable General Equilibrium (CGE) modeling has been placed on calibrating the baseline dynamics to forecasts. This paper suggests the formal method to calibrate all exogenous parameters of the Dynamic Applied Regional Trade (DART) model to forecasts from the the World Energy Outlook 2018 report. First, we determine the exogenous parameters (inputs) and forecasts (outputs) for the calibration procedure. Then we use the metamodeling method to generate surrogate models for the DART model. In the next step, we implement the Maximum A Posterior (MAP) method to estimate the exogenous parameters that are used to calibrate the baseline dynamics. Finally, we run the simulation with the estimates to test the performance.

Contents

1	Introduction	3
2	DART Description and Parameters Introduction	3
3	Bayesian Estimation	9
4	Explanation of Metamodeling: Concepts, Advantages and Process	11
4.1	Concepts	11
4.2	Advantages	12
4.3	Process	12
4.3.1	Determining Input(s) and Output(s)	12
4.3.2	Design of Experiments (Sampling Methods)	13
4.3.3	Metamodels (Approximation Methods)	15
4.3.4	Running Simulation and Collecting Data	16
4.3.5	Estimation and Validation	16
4.3.6	Using Metamodel for Intended Purposes	17
5	Results and Conclusions	18
5.1	Validation Results	18
5.2	Bayesian Estimation Results	18
5.3	Simulation Results with the Estimated Parameters	18

Acronyms

CGE Computable General Equilibrium.

DART Dynamic Applied Regional Trade.

DOE Design of Experiments.

LHS Latin Hypercube Sample.

MAP Maximum A Posterior.

SRS Simple Random Sampling.

1 Introduction

Computable General Equilibrium (CGE) models are workhorse models in applied economic policy analysis and particularly play an important role in developing economic evidence and modeling climate and energy policies. As a member of this family, the DART model (Dynamic Applied Regional Trade Model) is a multi-regional, multi-sectoral recursive dynamic model of the world economy. It was designed to analyze long-term trends resulting from international climate, energy, agricultural and land-use policies with regard to the relevant interactions and feedback effects like trade effects and sectoral adjustments.

The goal of this paper is to offer a formal way to calibrate the exogenous parameters in the DART model to the dynamics. The idea comes from such a perspective that there are enormous combinations of these parameters which lead to corresponding dynamics. We assume that the dynamic that we want to match is one dynamic from the countless dynamics. Therefore, what we are trying to do is to estimate these parameters based on the data we have and the idea is to use the Maximum a Posterior (MAP) method from the Bayesian estimation. Furthermore, to apply to accomplish the task, we need to include the DART model into the MAP method and that is technically infeasible due to large computational times. Therefore, we plan to use the metamodeling method to generate surrogates of the DART and combine them with the MAP to do the estimation. Finally, once the estimation is finished, we will plug them into the DART and check the dynamics produced by the estimated parameters.

The paper is arranged such that we start with a short model description of DART which includes a discussion on the choice of parameters included in the study. This is followed by a description of the Bayesian estimation technique. Subsequently, we have the meta modelling approach described in detail which is then followed by preliminary results and conclusions.

2 DART Description and Parameters Introduction

The DART (Dynamic Applied Regional Trade) model is a multi-region, multi-sector recursive dynamic CGE-model of the world economy. Labor, capital and land are the primary production factors. The economy in each region is modeled as a competitive economy with flexible output and factor prices and market clearing condition. Labor is homogenous and domestically mobile across sectors but not across countries. Capital is modeled with the vintage capital approach i.e. only the new capital is freely allocated across different sectors and the capital stock

from the previous period becomes sector bound in the subsequent periods. Land is an input only in the agricultural sectors. The regions are connected by bilateral trade flows. The DART model is a recursive dynamic model and it solves for the equilibrium solution in every year.

Producers minimize the costs of production for output production. The production functions are defined by constant elasticity of substitution (CES) functions that through different nesting structures allow for substitution between labor, capital, energy and non-energy intermediates. A general nesting structure in DART is shown in Figure 1.

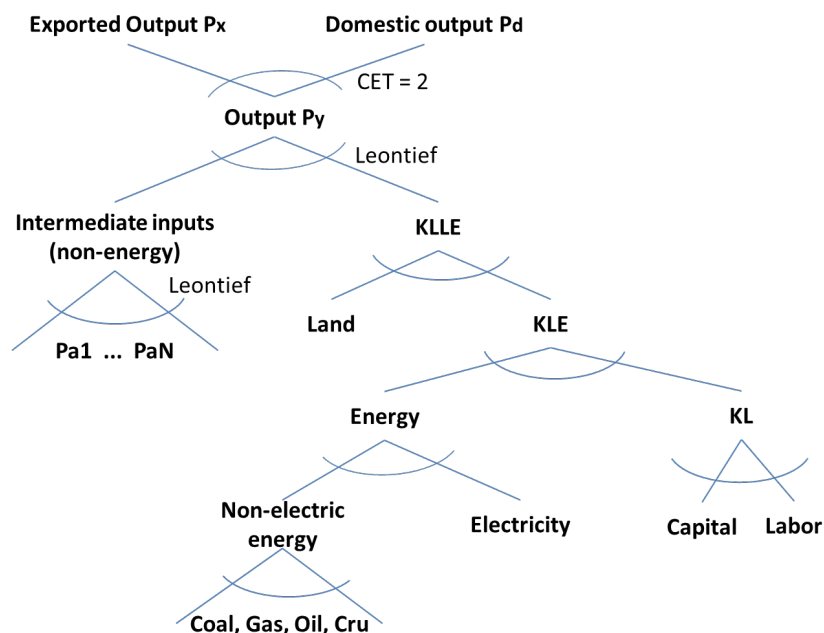


Figure 1: A General Nesting Structure in DART

The consumer receives all the factor incomes. The consumption demand is modeled as a linear expenditure system (LES) wherein the consumption is differentiated between subsistence consumption and surplus consumption. The factor markets are perfectly competitive and full employment of all factors is assumed. The dynamics of the model are driven by the rate of productivity growth, the savings rate, the rate of change of the workforce, and the change in human capital. For a more detailed description of the DART model, see Klepper et al. 2003.

In this paper we use the DART-CLIM version which has a detailed description of the energy sector with the inclusion of renewable technologies namely solar PV, wind and hydropower, and nuclear for electricity production. The model is calibrated to the GTAP 9 and GTAP power database with base year in 2011. The

FRA	France
DEU	Germany
GBR	United Kingdom, Ireland
BLX	Belgium, Netherlands, Luxembourg
SCA	Denmark, Finland, Sweden
MED	Mediterranean region
EEU	Eastern European Countries
REU	Rest of Europe
USA	United States of America
CAN	Canada
RAXB	Rest Annex B (Japan, Australia and New Zealand)
RUS	Russia
FSU	Former Soviet Union
CPA	China and Hong Kong
IND	India
LAM	Latin America
PAS	Pacific Asia
MEA	Middle East and North America
AFR	Sub Saharan Africa
BRA	Brazil

Table 1: Regions in DART Model

horizon is until 2030. The regions and sectors in our model are defined in Table 1 and Table 2.

In the DART dynamic baseline calibration we specially consider the renewable technology pathways and so we will briefly take a closer look at how these sectors are modelled. The nesting for the sectors is given in Figure 2.

Globally the renewable electricity sectors have undergone tremendous growth since 2011 as can be calculated from the International Energy Agency database. This spur in growth has led to a steady drop in their costs in the past years. To correctly define the baseline, this increase in the renewable technologies should be included in the baseline. To implement this decrease in input cost, in the DART model we introduced a learning curve model in the renewable electricity sectors. This is one of the most commonly used model forms and it associates the unit cost of electricity technology to its generation. The learning curve model intuitively also makes sense since it can be interpreted in the manner that typically learning by doing causes drop in costs of technologies. The classic characteristic parameter is called the learning factor and it can be understood as the fractional reduction in cost for each doubling of cumulative production. (Rubin et al., 2015)

COL	Coal
GAS	Natural Gas
CRU	Crude Oil
OIL	Petroleum and coal products
GRN	Grains
OSD	Oil seeds
CRO	Rest of crops
ANI	Animal products
RAGR	Rest of agriculture
CRP	Chemical products
ETS	Energy intensive products
MOB	Mobility
OLI	Other light industries
OHI	Other heavy industries
SVCS	Services
CGD	Savings good
EGAS	Gas based electricity
ECOAL	Coal based electricity
EHYDRO	Hydro based electricity
ENUCLEAR	Nuclear based electricity

Table 2: Sectors in DART Model

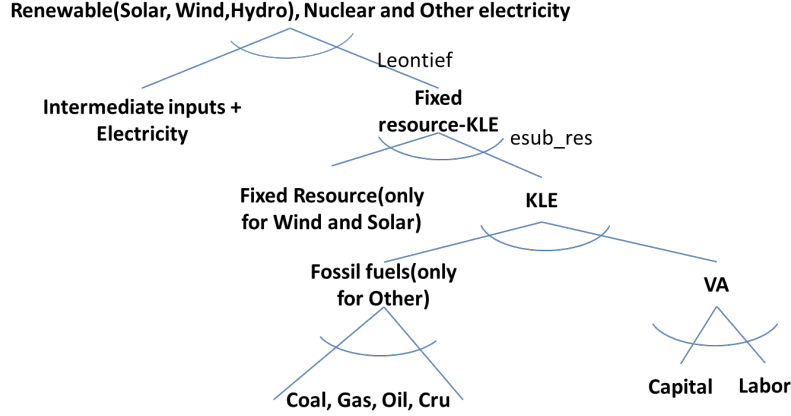


Figure 2: The Nesting for Sectors.

Previously, this has been implemented in the dynamic version of DART and a detailed description can be found in Weitzel 2010.

Mathematically, we use the following transformation to change the progress rates into learning rates for the technologies. It can be seen that the learning factor is calculated in each period since the ratio of the cumulative production relative to base year production changes in every period and so the more production of electricity from certain technologies the more learning is seen in these sectors.

$$learningfactor^t = \frac{\sum_r Y_r^t \frac{\log(progressrate)}{\log(2)}}{\sum_r Y_r^{2011}} \quad (1)$$

In the renewable technologies nest we see a fixed resource nested with KLE aggregate. This production structure for the renewables has also been used in other CGE models (Paltsev et al., 2005). The role of the fixed factor is to ensure that there are no flip-flop movements in the production of the renewables. This is important since we have a homogenous price of electricity irrespective of the production technology used and so the fixed resource is necessary.

The fixed resource for Solar and Wind is a share of the existing capital in that sector. Note that we do not use the fixed factor for nuclear and hydro because we assume that the production values for these technologies are principally policy driven. Using the World Energy Outlook 2018 report we exogenously define a growth rate for these technologies based on the forecast and thus they are not endogenously determined in our model.

After the model is set up, to conduct any policy analysis we have to define a robust baseline. The baseline can be defined in different ways wherein the researchers can choose which model outcomes they want to match with projections. For example, for a research question which is interested in examining the effect of

a climate policy on regional and global emissions it is crucial to have the model on a reasonable emission pathway in the baseline dynamics. And in addition to this to simultaneously also examine the macroeconomic impacts of the policy shock one might also consider having a reasonable regional GDP growth pathway or investment pathway.

This is exactly where the niche of the method we propose lies. It is typically possible to adjust more than one exogenous input parameters to get to the target GDP pathway, emission pathway or/and energy use pathway, whatever one seeks in their baseline. Moreover, since CGE models are over-identified there are several input parameter combinations that could put one this target path in the baseline. However, in dynamic CGE models with largely disaggregated regions and sectors like DART, manually conducting this exercise is extremely tedious and often one can never know whether we have chosen a ‘correct’ parameter combination for the baseline.

It is a fairly common practice to take the exogenous input parameters from other sources or models. However, the structure of models differs from each other and therefore direct adoption of these parameter values without carefully considering model differences is not the best method. Moreover, since the regional and sectoral disaggregation differs across models and also in econometric estimation studies several assumptions need to be made for sectors for whom these parameters are not estimated in the literature.

The motivation for this paper comes from these characteristics of calibrating a CGE model to a dynamic baseline.

In this paper, we introduce Bayesian and metamodeling techniques to calibrate the regional GDP pathway and regional electricity production of renewable technologies (Solar PV and Wind) until 2030 in the DART model. Since we aim to assess the impact of climate policies in the renewable technologies we have chosen these output parameters as of interest for us. The method allows for the number of calibration parameters to be increased or decreased based on the needs of the research question. The target data for the renewable electricity production for DART regions is constructed from World Energy Outlook 2018 report.

Since the dynamics of the model is driven by change in human capital, technical progress, savings rate we choose these parameters to calibrate the GDP pathway. For the electricity pathways we use the learning rate, elasticity between the fixed resource and KLE nest and the share of fixed resource for calibration. It is fairly intuitive that the two sets of calibration parameters are not independent and one has an effect on the others output variable which is where the challenge of simultaneously calibrating the baseline to these two pathways for every region in the model lies.

Changes in labor force are driven by employment growth rate and technical

progress. This change in labor force is one of the factors that drive annual changes in production. Similarly, the technical progress parameter also leads to increase in resource efficiency and so affects the production values from sectors. DART is a savings driven model and so it is another parameter that drives the model dynamics since it determines the value of investments in the different years which is used as capital in the next period. All these three parameters affect the production from different sectors which directly affect the regional GDP and so we have chosen them for our study.

By using our proposed methodology we are able to estimate the above mentioned exogenous parameter values of our calibration parameters which are tailor-made according to the structural features of DART.

3 Bayesian Estimation

The foundation of Bayesian Estimation is the well-known Bayes's rule (Bayes, 1991).

In probability theory and statistics, Bayes's rule describes the probability of an event, based on prior knowledge of conditions that might be related to the event. Mathematically, it can be represented by the following formular:

$$Pr(A|B) = \frac{Pr(B|A)Pr(A)}{Pr(B)} \quad (2)$$

where A and B are events and $Pr(B) \neq 0$. $Pr(A)$ and $Pr(B)$ are the probabilities of A and B occurring independently of each other. $Pr(B|A)$ is conditional probability of event B occurring given that A is true while $Pr(A|B)$ is the other way round.

In Bayesian estimation, this idea is generalized and applied to probability distribution functions of parameters instead of probabilities of events:

$$Pr(\theta|X) = \frac{Pr(X|\theta)Pr(\theta)}{Pr(X)} \quad (3)$$

where $Pr(\theta|X)$ represents the posterior distribution of θ given data X , $Pr(X|\theta)$ is the sampling density for X given θ , $Pr(\theta)$ is the prior distribution of parameters θ and $Pr(X)$ refers to the marginal probability of X .

For the sake of further derivation and following Lynch 2007, equation 3 can be transformed into a simplified form:

$$Pr(\theta|X) \propto \mathcal{L}(X|\theta)Pr(\theta) \quad (4)$$

where $\mathcal{L}(X|\theta)$ is the likelihood of X given θ .

Therefore the goal is to find estimates of parameters θ such that they maximize the posterior density $Pr(\theta|X)$, which is also known as the Maximum A Posterior (MAP) method. Mathematically, θ is estimated by:

$$\begin{aligned}\hat{\theta}_{MAP} &= \underset{\theta}{argmax} \mathcal{L}(X|\theta)Pr(\theta) \\ &= \underset{\theta}{argmax} \prod_{x_i \in X} Pr(x_i|\theta)Pr(\theta) \\ &\propto \underset{\theta}{argmax} \sum_{x_i \in X} (\log Pr(x_i|\theta)) + \log Pr(\theta)\end{aligned}\tag{5}$$

Let's assume that there is a forecast gdp growth rate gdp that we want to match, namely, we aim at estimating θ to make the DART model produce its own gdp growth rate GDP that is as close to gdp as possible. So we assume that there is a random deviation ϵ between gdp and GDP : $gdp = GDP(\theta) + \epsilon$. Following what we have derived, we can obtain estimates of θ by :

$$\begin{aligned}\hat{\theta}_{MAP} &= \underset{\theta}{argmax} \mathcal{L}(\epsilon|\theta)Pr(\theta) \\ &\propto \underset{\theta}{argmax} (\log Pr(\epsilon|\theta)) + \log Pr(\theta) \\ &s.t. gdp = GDP(\theta) + \epsilon\end{aligned}\tag{6}$$

Apart from the assumption that ϵ is normally distributed, we assume additionally that ϵ has a zero mean due to the fact that we want to match gdp and GDP , thus we have $\epsilon|\theta \sim \mathcal{N}(0, \sigma_\epsilon^2)$. Moreover, it is assumed that we have prior knowledge about the distribution of θ , $\theta \sim \mathcal{N}(\mu_\theta, \sigma_\theta^2)$, following the derivations in Heckelei and Mittelhammer 2008, we can get:

$$\begin{aligned}MAP_\theta &= \underset{\theta}{argmin} \left(\frac{\epsilon}{\sigma_\epsilon} \right)^2 + \left(\frac{\theta - \mu_\theta}{\sigma_\theta} \right)^2 \\ &s.t. gdp = GDP(\theta) + \epsilon\end{aligned}\tag{7}$$

Equivalently, we are estimating θ by minimizing the error between the forecast gdp growth rate and its counterpart of the DART model as well as taking into consideration the deviation of θ from its prior. Essentially, what we mean by “calibrating the exogenous parameters to the baseline dynamics” is “estimating the parameters such that the dynamics produced by DART match the forecasts” in its nature.

As stated before, in this paper, we narrow down our research scope to three outputs, gdp growth rate, electricity production from solar and electricity production from wind. Unfortunately, it is technically not feasible to combine the DART model directly with the MAP method because in the essence of this method, we are searching from enormous possible combinations of parameter values for the “one” that minimize the errors. Each potential combination needs to be simulated in order to provide data for the MAP method and this almost gives rise to an impossible mission. Therefore, we need simplified surrogates of the DART model which can be easily and fast incorporated into the MAP method and this is where the metamodeling method comes into the picture.

4 Explanation of Metamodeling: Concepts, Advantages and Process

4.1 Concepts

The metamodeling method has a long history of assisting researchers carry out computer experiments and develop simulation models. It aims at approximating and understanding the input-output relationship of the simulation models. The idea and term of metamodeling were originated by Robert Blanning (Blanning, 1975) and are developed and popularized by Jack Kleijnen (Kleijnen, 1975). Simulation model is the first level abstraction of the reality and we can think of metamodeling as the second level abstraction to deal with some problems. For a short review of the history and early works of metamodeling, see Barton 2015.

Before going to more details, we want to make two clarifications on the terminology issue in this paper. In the fields such as engineering, design optimization and so forth,, metamodeling is sometimes also named as surrogate modeling, design and analysis of computer experiments, response surface modeling and etc, these terms refer to the same concept. In metamodeling, we consider the simulation model as a black box and try to use a function, say f_{meta} , to approximate the input-output behavior. The function f_{meta} is called metamodel in the context and constructed by two parts, design parameters x and responses y . The term design parameters and responses refer to the inputs and outputs of the simulation model respectively. Throughout this paper, we use these terms interchangeably.

To grasp the fundamental idea of metamodeling, let’s assume that we have a simulation model which models such a relationship:

$$y = SIM(x), \tag{8}$$

where x and y represent the simulation inputs and outputs that are of great interest to us.

The metamodeling method has the purpose of approximating this relationship by means of a more simplified model, say f_{meta} :

$$\hat{y} = f_{meta}(x), \quad (9)$$

where $\hat{y}$ refers to the metamodel predictions.

4.2 Advantages

Roughly speaking, the metamodeling method has two advantages. Firstly, it generates simplified explicit metamodels which are much more user-friendly than simulation models because they provide more insights into what is actually happening inside the simulation models and understand the input-output behavior better. Besides, the metamodels are computationally inexpensive to construct and evaluate (Kleijnen and Sargent, 2000; Kleijnen, 2015). Secondly, if the metamodels are successfully validated, they can be used as the surrogate of the simulation model which brings us great convenience:

1. Using metamodels as surrogate of the simulation model means that instead of running the simulation model repeatedly due to changing parameter values, we can simply calculate the output values. This could save both time and expenses.
2. As we know, it could get quite complicated if we want to apply the simulation model to other research purposes, such as optimization, sensitivity analysis and etc. However, these tasks become much more handy if we make use of metamodels instead.

4.3 Process

In the literature, there are multiple versions of metamodeling processes that are recommended by various authors such as Russel Barton and Jack Kleijnen (Barton, 2015; Kleijnen and Sargent, 2000). We modified the version of Russel Barton slightly that is given in Table 3. There are 5 important steps in the process with several crucial components of the metamodeling method. We will introduce them while we go through the whole process.

4.3.1 Determining Input(s) and Output(s)

In section 2, we have introduced the inputs and outputs that are of interests to us in this paper.

Step	Task
1	Determining Input(s) and Output(s)
2	Selecting Metamodel types and DOE
3	Running Simulation and Collecting Data
4	Estimating and Validating the Metamodel
5	Using Metamodel for Intended Purposes

Table 3: The Metamodeling Process

4.3.2 Design of Experiments (Sampling Methods)

Design of Experiments (DOE) refers to the sampling method that is used to efficiently generate a sample of the design parameters x . In the literature, there are generally two streams of design of experiments, the “classical” design and the “space-filling” design. The “classical” design methods tend to place the sample points repeatedly at the boundaries of the design space and a few sample points at the center of the design space for the sake of accounting for the random variation of the stochastic simulation system. However, as is argued by Jerome Sacks in the paper Sacks et al. 1989 that the property of deterministic simulation models makes the idea of dealing with randomness unnecessary and therefore he and other authors advocate the “space-filling” design which spreads the sample points across the design space to collect as much information as possible. Figure 3 gives us a visual impression of the two streams of designs.

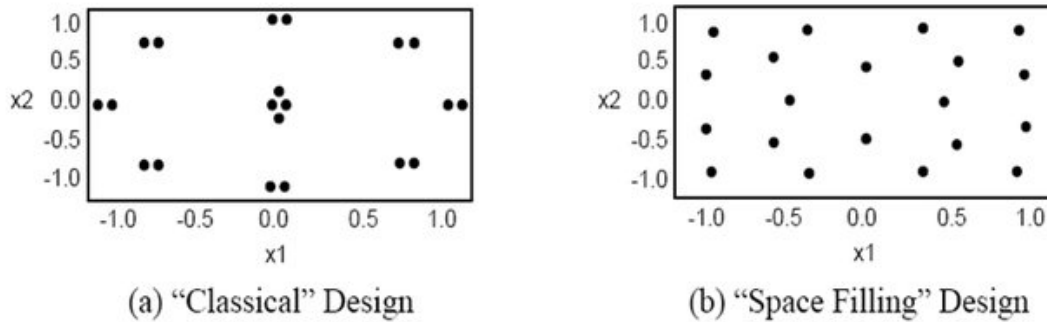


Figure 3: “classical” and “space-filling” Designs (Booker, 1998)

Both “classical” and “space-filling” designs contain several family members, for example, in the family of “space-filling” designs there are orthogonal arrays (Owen, 1992), Latin Hypercube Designs (McKay et al., 1979), Hammersley sequences

(Kalagnanam and Diwekar, 1997) and so forth. For more detailed reviews of DOE, we recommend you to have a look at the paper Dey et al. 2017.

In this paper, we apply the Latin Hypercube Sampling to generate the simulation sample. Latin Hypercube Sampling (LHS) is a statistical technique for generating a random sample from a multidimensional distribution. The sampling method is often applied to construct computer experiments.

In the context of statistical sampling, a square grid with sample points is termed as a Latin Square if (and only if) there is only one sample point in each row and each column (see Figure 4). The LHS is a generalization of this concept to an arbitrary number of dimensions, whereby each sample is the only one in each axis-aligned hyperplane containing it.

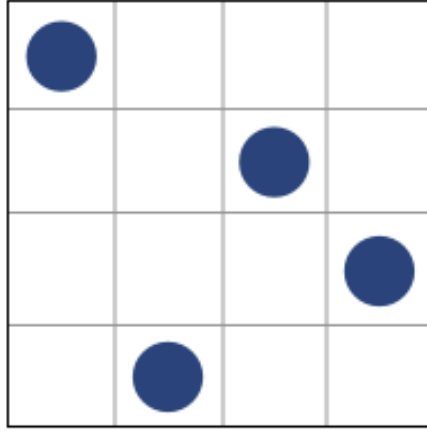


Figure 4: A Two-Dimensional Latin Hypercube Sample (Or a Latin Square)

Apart from the fact that “space-filling” designs work better with deterministic simulation models, several studies have shown that Simple Random Sampling (SRS) is not an efficient sampling strategy due to its less appropriate space-filling property. (Helton and Davis, 2003; Janssen, 2013; Rajabi and Ataie-Ashtiani, 2014) Moreover, Michael Stein and other authors (Stein, 1987; Florian, 1992) have pointed out that LHS is able to provide variance reduction of commonly used statistical parameters. Last but not the least, it is also demonstrated in the paper Raue et al. 2013 that LHS helps to produce more accurate estimates in comparison with SRS.

The LHS algorithm works as follows: let’s suppose that we have k design parameters and we would like to have a sample size n . Thus the Latin Hypercube algorithm divides the distribution of each design parameter x into n equally sized bins with the same probability. Then a point will be selected from each bin at random and this process will be applied to every design parameter. In the next step, the n points of the first design parameter will be randomly matched with the

n points of the second one and then this established pair will be matched again with the n points of the third design parameter. The process continues constantly and automatically until it completes the random pairing with the n points of the last design parameter and successfully form the $n - dimensional - pair$ which is our simulation sample. The standard LHS algorithm produces a sample without correlation among design parameters.

Furthermore, we want to mention that in the group of Latin Hypercube Sampling, there are various candidates supported by corresponding optimization criterion. We select the Genetic Latin Hypercube Sampling which draws a latin hypercube sample based on the idea of optimizing the sample with respect to the $S - Optimality$ criterion through a genetic type algorithm. $S - Optimality$ seeks to maximize the mean distance from each design point to all the other points in the design space, so the points are as spread-out as possible.

4.3.3 Metamodels (Approximation Methods)

The term metamodel refers to the functional form f_{meta} to approximate the input-output behavior of the simulation model. Possible candidates are polynomial metamodels, ordinary Kriging metamodels, radial basis functions and so forth. More introductions can be found in Dey et al. 2017; Simpson et al. 2001; Chen et al. 2006. In this paper, we give introductions of two popular metamodel types.

Polynomial Metamodels The polynomial metamodels were firstly developed by Box and Wilson in their paper Box and Wilson 1951 and then have been used in various applications. We can have a look at a first-order polynomial metamodel:

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \epsilon, \quad (10)$$

where the x_i , y and ϵ refer to the design parameters, simulation output and error term respectively. We can easily see that the polynomial metamodels are essentially equivalent to the multivariate regression equations in standard econometrics with one dissimilarity that we are now using simulation data instead of observational data to estimate the coefficients by means of the least squares regression:

$$\hat{\beta} = [X'X]^{-1}X'y \quad (11)$$

Ordinary Kriging The Kriging metamodels are originally invented in geostatistics (Cressie and Chan, 1989; Cressie, 1993), an ordinary Kriging metamodel takes

the following form:

$$y(x) = \mu + M(x), \quad (12)$$

where μ is the constant mean $E[y(x)]$ in the given $k - dimensional$ design space, and $M(x)$ is assumed to be a realization of a stochastic process and spatial correlation function given by:

$$Cov[M(x_i), M(x_j)] = \sigma^2 R(x_i, x_j) \quad (13)$$

where R refers to a certain correlation function which calculates the correlation based on the spatial distance between the points.

OK uses a linear predictor which predicts the value of a new point by means of a weighted sum of the values of old points and the optimal weights are determined by the best linear unbiased predictor (BLUP) criterion which minimizes the mean squared error (MSE) of the predictor.

Simply speaking, the Kriging metamodel assumes that the value of a new point is affected by the values of each old point and the closer an old point is to the new point, the larger the value of the old point affects the value of the new point, namely, the larger the weight of the old point is.

The maximum likelihood estimation method is used to obtain the estimates of corresponding parameters. For detailed explanation, we refer to Kleijnen 2015.

4.3.4 Running Simulation and Collecting Data

The number of the samples was defined such that we have sufficient number of simulations to be able to employ the meta-modeling techniques. For this application, we have $n = 1000$ simulations.

4.3.5 Estimation and Validation

After collecting the data, we can begin with estimation and validation. Estimation depends on the metamodel we choose and we will talk about it in the application part.

Validation is a crucial step in building metamodels because it is necessary for us to evaluate the performance of the metamodel and decide if they are qualified. To decide whether the performance is satisfactory or not, we need criteria. In the literature, authors have suggested multiple statistics and in this paper, we tend to use two popular ones, namely, the root mean squared error (RMSE) and the maximum absolute relative error (MARE):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (14)$$

$$MRAE = \max_i \left| \frac{y_i - \hat{y}_i}{y_i} \right|, \quad (15)$$

where y_i refers to the real simulation output and $\hat{y}_i$ refers to the predicted output from the metamodel. Intuitively, *RMSE* provides good evaluation of the “global” error over the region of interest and *MARE* offers a good evaluation of the “local” error by assessing the worst error within the region of interest (Simpson et al., 2001). To determine whether the values of them are good is always tricky because it is highly dependent on our view about how accurate we want the metamodel to be as well as on the research purposes. For example, to what extent do we care about the *MARE* as it measures the worst error with the region of interest, so if the research purpose is to pursue an extremely accurate metamodel without unacceptable errors, then *MARE* must be a deciding factor in drawing conclusions. If we care more about the global performance and have the tendency to live with certain bad predictions, then we will give more tolerance to *MARE* and pay more attention to *RMSE*. In addition, we provide the adjusted R^2 to serve as an extra reference.

Furthermore, we split the dataset into two sub-samples which we refer to training and test samples respectively. We use the training sample to estimate the metamodel and make predictions in terms of the test sample. Then we can compute the aforementioned two statistics in order to measure the performance of our metamodel, or, using the metamodeling jargon, validate our metamodel. To make the validation process robust, we bootstrap this process and evaluate the performance based on the averaged values of the statistics.

Basically, we can draw conclusions regarding the validity of our metamodel based on the validation results.

4.3.6 Using Metamodel for Intended Purposes

At the end of section 3, we argued that we are in need of surrogates of the DART model to make the MAP work. We plan to generate metamodels for each output with respect to each region to approximate the corresponding input-output relationship of the DART model and implement them in the MAP.

5 Results and Conclusions

5.1 Validation Results

5.2 Bayesian Estimation Results

5.3 Simulation Results with the Estimated Parameters

References

- Barton, R. R. (2015). Tutorial: simulation metamodeling. In *Proceedings of the 2015 Winter Simulation Conference*, pages 1765–1779. IEEE Press.
- Bayes, T. (1991). An essay towards solving a problem in the doctrine of chances. 1763. *MD computing: computers in medical practice*, 8(3):157.
- Blanning, R. W. (1975). The construction and implementation of metamodels. *Simulation*, 24(6):177–184.
- Booker, A. (1998). Design and analysis of computer experiments. In *7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, page 4757.
- Box, G. E. P. and Wilson, K. B. (1951). On the experimental attainment of optimum conditions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 13(1):1–45.
- Chen, V. C., Tsui, K.-L., Barton, R. R., and Meckesheimer, M. (2006). A review on design, modeling and applications of computer experiments. *IIE transactions*, 38(4):273–291.
- Cressie, N. and Chan, N. H. (1989). Spatial modeling of regional variables. *Journal of the American Statistical Association*, 84(406):393–401.
- Cressie, N. A. C. (1993). *Statistics for Spatial Data*. John Wiley & Sons, Inc.
- Dey, S., Mukhopadhyay, T., and Adhikari, S. (2017). Metamodel based high-fidelity stochastic analysis of composite laminates: A concise review with critical comparative assessment. *Composite Structures*, 171:227–250.
- Florian, A. (1992). An efficient sampling scheme: updated latin hypercube sampling. *Probabilistic engineering mechanics*, 7(2):123–130.
- Heckelei, T. and Mittelhammer, R. (2008). A bayesian alternative to generalized cross entropy solutions for underdetermined econometric models. Discussion Paper 2, Institute for Food and Resource Economics University of Bonn.
- Helton, J. C. and Davis, F. J. (2003). Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Engineering & System Safety*, 81(1):23–69.

- Janssen, H. (2013). Monte-carlo based uncertainty analysis: Sampling efficiency and sampling convergence. *Reliability Engineering & System Safety*, 109:123–132.
- Kalagnanam, J. R. and Diwekar, U. M. (1997). An efficient sampling technique for off-line quality control. *Technometrics*, 39(3):308–319.
- Kleijnen, J. P. (1975). A comment on Blanning’s “Metamodel for sensitivity analysis: the regression metamodel in simulation”. *Interfaces*, 5(3):21–23.
- Kleijnen, J. P. (2015). Design and analysis of simulation experiments. In *International Workshop on Simulation*, pages 3–22. Springer.
- Kleijnen, J. P. and Sargent, R. G. (2000). A methodology for fitting and validating metamodels in simulation. *European Journal of Operational Research*, 120(1):14–29.
- Klepper, G., Peterson, S., et al. (2003). On the robustness of marginal abatement cost curves: the influence of world energy prices. Technical report, Kiel Working Paper.
- Lynch, S. M., editor (2007). *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*. Springer New York.
- McKay, M. D., Beckman, R. J., and Conover, W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245.
- Owen, A. B. (1992). Orthogonal arrays for computer experiments, integration and visualization. *Statistica Sinica*, pages 439–452.
- Paltsev, S., Reilly, J. M., Jacoby, H. D., Eckaus, R. S., McFarland, J. R., Sarofim, M. C., Asadoorian, M. O., and Babiker, M. H. (2005). The mit emissions prediction and policy analysis (eppa) model: version 4. Technical report, MIT Joint Program on the Science and Policy of Global Change.
- Rajabi, M. M. and Ataie-Ashtiani, B. (2014). Sampling efficiency in monte carlo based uncertainty propagation strategies: application in seawater intrusion simulations. *Advances in Water Resources*, 67:46–64.
- Raue, A., Schilling, M., Bachmann, J., Matteson, A., Schelke, M., Kaschek, D., Hug, S., Kreutz, C., Harms, B. D., Theis, F. J., et al. (2013). Lessons learned from quantitative dynamical modeling in systems biology. *PloS one*, 8(9):e74335.

- Rubin, E. S., Azevedo, I. M., Jaramillo, P., and Yeh, S. (2015). A review of learning rates for electricity supply technologies. *Energy Policy*, 86:198–218.
- Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989). Design and analysis of computer experiments. *Statistical science*, pages 409–423.
- Simpson, T. W., Lin, D. K., and Chen, W. (2001). Sampling strategies for computer experiments: design and analysis. *International Journal of Reliability and Applications*, 2(3):209–240.
- Stein, M. (1987). Large sample properties of simulations using latin hypercube sampling. *Technometrics*, 29(2):143–151.
- Weitzel, M. (2010). Including renewable electricity generation and ccs into the dart model.