TECHNO-ECONOMIC OPTIMIZATION OF PHOTOVOLTAIC PLANT LAYOUT BY USING DESIGN OF EXPERIMENTS TECHNIQUES

Sascha Nadja Ringlstetter[1,2], Lukas Haack[1], Fritz Dildey[2], Lutz Sommer[1], Richard Meyer[1]

[1] Suntrace GmbH, Brandstwiete 46, 20457 Hamburg, Germany, +49 40 767 96 38 0, nadja.ringlstetter@suntrace.de, lukas.haack@suntrace.de, lutz.sommer@suntrace.de, richard.meyer@suntrace.de
[2] Hamburg University of Applied Sciences (HAW), Ulmenliet 20, 21033 Hamburg, +49 40 428 756244, f.dildey@ls.haw-hamburg.de

ABSTRACT: The conceptual design of utility scale photovoltaic power plants is commonly realized by using commercial or non-commercial simulation software tools for energy yield calculations.
Systematic techno-economic optimization by variation of several input variables with a reasonable step size can lead to a very large number of simulation runs. Also more detailed simulation algorithms increase the computational time.
Performance approximation of different plant layouts by applying Design of Experiments (DOE) methods and multiple linear regression models help to reduce required simulation cases and to identify the optimal plant layout more efficiently. Instead of step by step variation of all input parameters, only specific parameter configurations (design points) are simulated, which provides sufficient information for developing accurate regression models for approximation of the respective annual energy yield and Levelized Cost of Electricity (LCOE). The LCOE is used as measure for comparison of possible variants and identification of the optimal plant configuration.
High accuracy of the regression models could be achieved and demonstrated in a case study. With only 95 simulation runs for determining the approximation function, LCOEs of several thousands different layouts have been calculated with an uncertainty lower than 3%. Computation time has been reduced by a factor of five compared to a direct reference calculation, which corresponds to saving several hours of worktime.

Keywords: PV system, Simulation, Utilities

1 MOTIVATION

During the techno-economic optimization procedure for utility scale PV plants, several different design configurations are examined. When simulating the PV plant's energy production, by using detailed simulation software or applying high time resolutions, the calculation time increases significantly. The time required for the optimization procedure also depends on the computing power. Minimizing the number of required simulation runs and estimating all remaining possible plant configurations with satisfying accuracy can reduce the time and hardware effort. A regression model can be used to predict different output variables, such as annual energy yield, Performance Ratio (PR) or directly Levelized Cost of Electricity (LCOE). Furthermore, influence of single input parameters on the output variables e.g. over dimensioning of PV array, row distance or other factors can be analysed systematically.

This fast approximation allows analysing a PV project in conceptual design state systematically and prevents biasing the result of the techno-economic optimization by too strict limitation of the input variables.

2 METHODOLOGY

The proposed approach is based on Design of Experiments (DOE) techniques, which aim to describe a system's behaviour with as few simulation runs as possible by gaining most information for system approximation via regression models. First, a set of plant configurations (design points) is determined and simulated. Based on response results for the design point a suitable regression model is configured.

2.1 Selection of factors and responses

The first step is defining the system onto which the DOE method shall be applied. In this case, the system is a utility scale PV plant, or more accurate the model of the plant used by the PV simulation software. The selected factors and responses form the basis for DOE design. Plant parameters, which are not varied due to either software restrictions or minor importance regarding the impact on the response, are not considered in the design and are held constant. Whereas factors affecting the response significantly are considered in the design and varied according to selected design points during the simulation. A 50 MW PV power plant in sunbelt region was assessed in a case study with variation of four factors as presented in the following table I. As response values, the annual energy yield and Levelized Cost of Electricity (LCOE) have been chosen.

Table I: Factors varied in 50 MW plant case study (module materials mono-Si: monocrystalline silicon, mc-Si: multicrystalline silicon, CdTe: Cadmium Telluride; sheds width: 4 m for mono- and multicrystalline modules, 2.4 m for thin film modules)

	Type	Min	Max
Sheds tilt	continuous	10°	40°
Sheds distance (pitch)	continuous	4 m	14 m
DC/AC sizing ratio	continuous	0.8	1.6
Module technology	categorical		mono-Si/ mc-Si/ CdTe/

2.2 Selection of design type

The selection of an appropriate experimental design depends, among others, on expected response characteristic. When high order models are required for appropriate description of response characteristics, then more design points are needed to develop respective regression models. On the other hand, the maximum number of design points is limited by the affordable time and effort, also in this case of a computer experiment. The simplest experimental design is the full factorial design, in which all combinations of input parameters, each varied with a certain step size, are assessed. The resulting design points are uniformly distributed over the multi-dimensional design space formed by the design factors. Experimental designs suitable for computer simulations are random and space-filling designs, the most popular being Monte-Carlo design and Latin Square (Hypercube) design [1][2]. In this work, D-optimal and I-optimal designs have been applied due to their ability to generate designs with a particularly small number of design points, and due to literature recommendations indicating their suitability for computer experiments. The underlying principle of D-optimal designs is minimizing the determinant of $(X'X)^{-1}$ (with X being the factor matrix), which is equal to minimizing the volume of the joint confidence region on the regression coefficient vector. In I-optimal designs (also called Q-, V- or IV-optimal, standing for integrated variance), the average integrated prediction variance over the design space is minimized. [1][2].

2.3 Realization of design points

In order to obtain response values for each parameter configuration specified in the DOE design, the annual energy yield has been calculated with the commercial PV simulation software PVsyst (version 6.3.9). The time needed for each run is mainly influenced by the working principle and internal algorithms of the software, and by the plant configuration specified by the user (e.g. number of strings). In the conducted case study, each simulation run required approximately 30 seconds in batch mode.

The LCOE for each plant configuration has been determined using a company-intern cash-flow model based on standard financial assumptions.

2.4 Regression model function

After determining the results of PV generator design point simulations and financial modelling of LCOE, a multiple linear regression (MLR) is applied to determine an approximation model. In this case, least square method was used for generating polynomial model function also including interaction terms of input factors. The function represents a hyper plane in the k-dimensional design space of the k regressor variables and aims to approximate the true response of a system.

$$y = f(x_1, x_2, \ldots, x_k) + \varepsilon \tag{1}$$

Where:
$\boldsymbol{y}$ = Response
$\boldsymbol{x_i}$ = Coded variable (factor)
$\boldsymbol{\varepsilon}$ = Statistical error (deterministic experiments: bias)

The following example illustrates a second-order predictive model with $\boldsymbol{x_i x_j}$ being two-factor interaction terms of factors i and j ($i < j$) [3].

$$y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + \sum_{i=1}^{k} \sum_{j=1}^{k-1} \beta_{ij} x_i x_j + \sum_{j=1}^{k} \beta_{jj} x_j^2 + \varepsilon_i \tag{2}$$

After linearization of model terms, the model for prediction of response values $\hat{y}$ can be determined with least square regression.

$$X'Xb = X'y \Rightarrow b = (X'X)^{-1}X'y \tag{3}$$

$$\hat{y} = Xb = b_0 + \sum_{j=1}^{k} b_j x_{ij} \tag{4}$$

Where:
$\boldsymbol{X}$ = Factor matrix
$\boldsymbol{\hat{y}}$ = Predicted response
$\boldsymbol{b_i}$ = Regression coefficients

The errors between actual responses and predicted values are called residuals.

$$e = y - \hat{y} \tag{5}$$

A suitable model requires careful selection of relevant model terms and elimination of non-appropriate terms with the help of statistical measures and indicators. With regard to selection of models, backward and stepwise were proven to give better results compared to forward selection, which is in well agreement with literature recommendations [1][3]. A transformation of the response, usually a power or logarithmic transformation, may be advantageous by means of stabilizing the response variance, making its distribution closer to the normal distribution or improving the fit of the model to the design point values. They can be realized in the following form:

$$y^* = \begin{cases} y^\lambda, \ \lambda \neq 0 \\ ln(\lambda), \ \lambda = 0 \end{cases} \tag{5}$$

Where:
$\boldsymbol{y}$ = Response
$\boldsymbol{y^*}$ = Transformed response
$\boldsymbol{\lambda}$ = Parameter of transformation

The selection of base model order and model terms was evaluated with different selection criteria indicating the quality of the resulting predictive model. Visual tools were applied in order to identify discrepancies, such as normal probability plot, predicted vs. actual plot, residuals vs predicted and vs. factor setting plots. A box-cox plot was used as support tool for choosing an appropriate response transformation. Further measures for evaluating different combinations of model terms were

- Predicted Error Sum of Squares PRESS, an indicator for the prediction quality of the regression model,
- adjusted R^2, which is "normalized" with regard to number of design points and number of model terms,
- standard deviation of the model.

The regression function characteristics obtained with an I-optimal design based on 95 design points, meaning 95 simulations, are presented in table II. The assessment of the quality of obtained models with the help of these value was done by comparing them to other model term combinations determined with the same set of design point values or with the best fitting regression models for different numbers of design points.

Table II: Regression function characteristics for energy yield and LCOE as response values, both obtained with I-optimal design, 95 design points

	Annual energy yield	LCOE
Order of base model	quartic	quartic
Model term reduction	stepwise	backward
Response transformation	none	ln
Adjusted R2	0.99995	0.99996
PRESS	7.9E+12	9.4E-04
Standard deviation	1.6E+05	2.2E-03

2.5 Comparison with reference data

For final evaluation of quality of obtained predictive models, their response predictions were compared to a full factorial reference set covering the same design space as the predictive models that has been directly simulated and calculated. The steps and resulting number of simulations are given in table III.

Table III: Factors variation in full factorial reference data set, based on the case study (module materials mono-Si: monocrystalline silicon, mc-Si: multicrystalline silicon, CdTe: Cadmium Telluride)

	Range	Step size	No. of variations
For each technology:			
Tilt continuous	10°…40°	5°	7
Pitch	4m…14m	2m	6
DC/AC sizing	0.8…1.6	0.2	5
Sum:			210
Total sum (3 module types)			630

The predictive models obtained for different numbers of design points and with different types of designs (D- and I-optimal) were evaluated with respect to mean bias, standard deviation and maximum absolute residual of predicted responses compared to reference data.

3 RESULTS AND CONCLUSIONS

The maximum deviations between different predictive models and number of simulations are illustrated in figures 1 and 2. Both figures clearly indicate that a higher number of design points yields in smaller maximum bias. As both D- and I-optimal designs lead to comparably small deviations, both design types can be regarded as being suitable for this application with a sufficiently high number of design points. However, the quality of regression model is strongly correlated with selection of model terms as described in section 2.4.

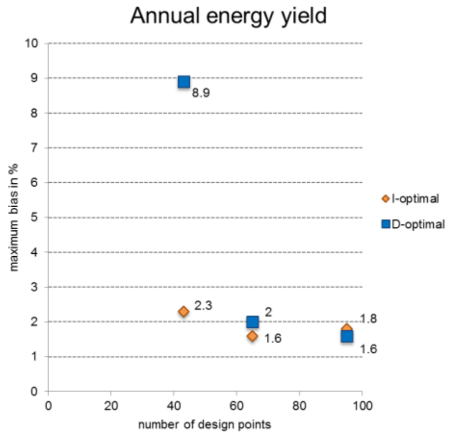

Figure 1: Maximum deviations of regression model predicted responses to reference values for annual energy yield

Figure 3 and 4 indicate that the regression models only slightly underestimate the responses for the I-optimal design based on 95 simulations. However, the model accuracy is definitely sufficient, as most absolute deviations are below 1 %.

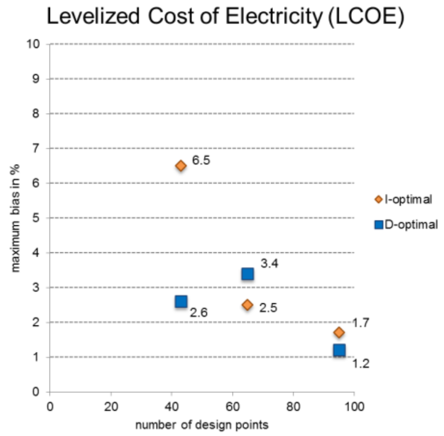

Figure 2: Maximum deviations of regression model predicted responses to reference values for LCOE

The uncertainty of the regression models can be taken into account by recalculating not only the plant layout with minimum LCOE as determined by the approximation function, but a larger number plant configurations out of the most favourable region of design space. When the top ten layouts are recalculated, it can be assumed that the real optimal layout is included in this set and thus can be determined.

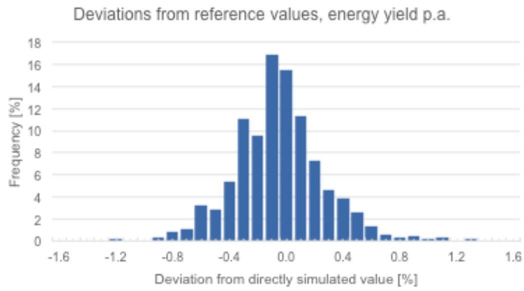

Figure 3: Distribution of deviations of predicted responses to reference values for annual energy yield, I-optimal design with 95 design points

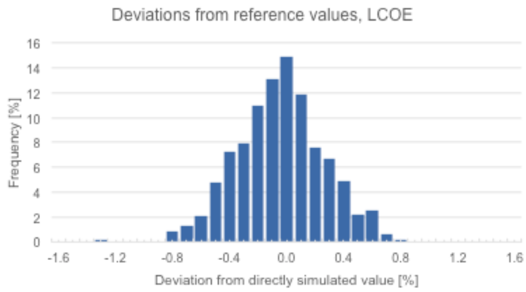

Figure 4: Distribution of deviations of predicted responses to reference values for LCOE, I-optimal design with 95 design points

Direct calculation of all 630 variants in the reference data set results in a time effort of more than 5 hours. The aim of the assessed approach was to reduce this time while maintaining a satisfying accuracy. With regard to the presented deviations, this target could be reached. The time needed for simulations was reduced by a factor of five to six. Some additional worktime is necessary for data handling and utilization of DOE software, on the other hand, a simple and fast possibility of analysing simulation results is provided by the regression function. One main advantage of obtaining a regression model function is the visualisation of predicted responses as three-dimensional surface plots and contour plots. With the help of the regression function it is possible to interpolate the response value for every plant layout within the design space. Visualisations indicating ranges of desirable or undesirable response values, which can be applied as decision support tool. Examples of such contour plots are provided by figures 5 to 8.

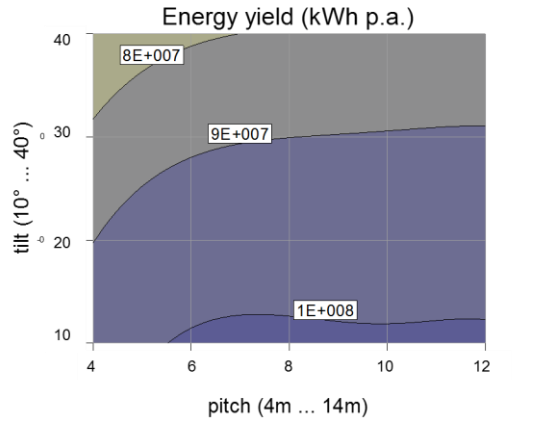

Figure 5: Interpolation of annual energy yield for different pitch and tilt values, at DC/AC sizing of 1.2 and mono-Si modules

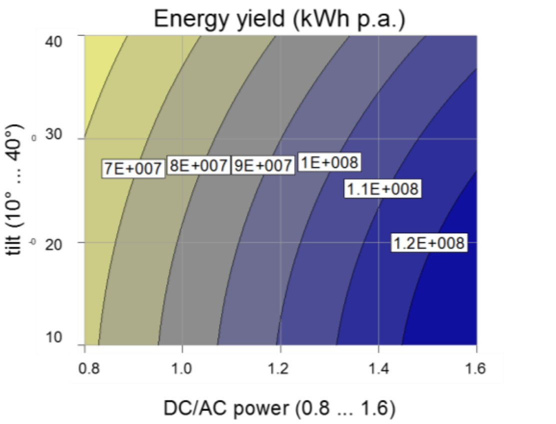

Figure 6: Interpolation of annual energy yield for different DC/AC sizing and tilt values, at pitch of 8m and mono-Si modules

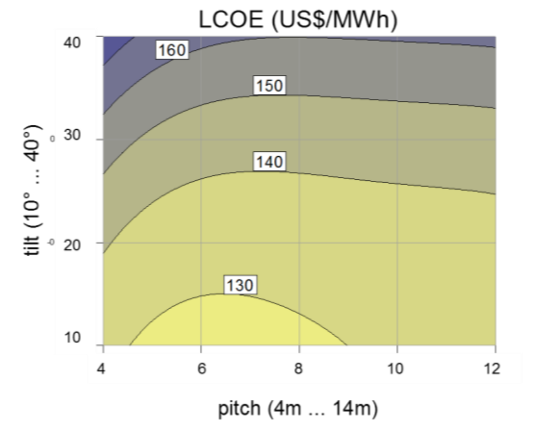

Figure 7: Interpolation of LCOE for different pitch and tilt values, at DC/AC sizing of 1.2 and mono-Si modules

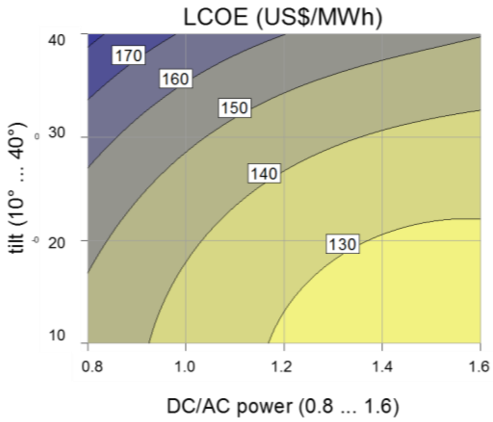

Figure 8: Interpolation of LCOE for different DC/AC sizing and tilt values, at pitch of 8m and monocrystalline modules

As illustrated by the contour plots, in this case study a small tilt angle, medium pitch and large DC/AC power sizing ratio is desirable in terms of minimized cost of electricity. All results are exemplary, as based on the applied technical and financial assumptions, such as site-specific solar resource, PV module characteristics and land costs. Especially differences between maximum energy yield and minimum LCOE are influenced by financial assumptions.

SUMMARY

The proposed method of applying DOE techniques for techno-economic layout optimization of utility scale photovoltaic plants shows high potential regarding saving of time and accuracy and reliability of results. D- and I-optimal designs with 95 design points resulted in standard deviations of 0.5% and 0.3% respectively and 1.7% and 1.2% of maximum absolute bias. An I-optimal design composed of 95 simulations and resulting in a polynomial regression function of fourth degree in terms of accuracy is most suitable for the examined case study. Time effort was reduced from more than five hours of simulation time to less then one hour, with some additional time for DOE software handling and data processing. However, the approach is based technical and financial assumptions. Knowledge of the underlying statistical methods is required for selecting the regression model terms, which significantly influence the quality of the predictive model.

REFERENCES

[1] Montgomery, D. C. (2009). *Design and Analysis of Experiments* (7th ed.). Singapore: Wiley.

[2] Siebertz, K., van Bebber, D., & Hochkirchen, T. (2010). *Statistische Versuchsplanung : Design of Experiments (DoE)*. Berlin, Germany: Springer.

[3] Myers, R. H., & Montgomery, D. C. (2002). *Response Surface Methodology : Process and Product Optimization Using Designed Experiments* (2nd ed.). New York, USA: Wiley.