PROCEEDINGS OF ECOS 2023 - THE 36TH INTERNATIONAL CONFERENCE ON
EFFICIENCY, COST, OPTIMIZATION, SIMULATION AND ENVIRONMENTAL IMPACT OF ENERGY SYSTEMS
25-30 JUNE, 2023, LAS PALMAS DE GRAN CANARIA, SPAIN

Optimization-based energy system planning under long-term uncertainty: Rapid Monte Carlo analysis using linear regression

Hagen Seele[a,b], Jan Tautorus[a,c], Christiane Reinert[a,d], Niklas von der Aßen[a,e]

[a] Institute of Technical Thermodynamics, RWTH Aachen University, 52062 Aachen, Germany
[b] hagen.seele@ltt.rwth-aachen.de
[c] jan.tautorus@rwth-aachen.de
[d] christiane.reinert@ltt.rwth-aachen.de
[e] niklas.vonderassen@ltt.rwth-aachen.de, CA

Abstract:

Decision-making for distributed energy systems (DES) is subject to significant uncertainties. Therefore, assuming perfect foresight for long-term system planning might result in suboptimal decisions. Long time horizons result in a variety of possible scenarios. One way of considering uncertainty in DES design is Monte Carlo analysis (MCA). However, MCA suffers from the computational burden of repeatedly evaluating energy system models. Furthermore, MCA is sensitive to distribution assumptions of uncertain parameters. In this work, we combine linear regression and systematic uncertainty modeling in MCA to consider uncertainty effectively.

We propose a method for comparing different DES designs regarding total annualized cost (TAC) while reflecting uncertain parameters. We model uncertainties by introducing a small number of representative factors that scale reference parameters. We distinguish between constant and long-term uncertainties increasing over time. We use mixed-integer linear programs (MILPs) to minimize yearly costs of DES. We solve the burdensome MILPs only for Latin Hypercube samples to parameterize linear surrogate models. We use the obtained linear surrogate models during MCA to accelerate the computation of the TAC.

We apply our method to a case study adapted from the literature to compare promising DES designs while considering multiple sources of uncertainty. We compute the TAC of the DES designs for thousands of long-term scenarios and identify the design, which results most frequently in the lowest costs. We show that using linear regression can reduce the computational time by more than 99 %, while maintaining a high accuracy measured by the goodness of fit of the linear regression.

Keywords:

Distributed energy systems; Long-term scenarios; MILP; Multiple uncertainties; Surrogate modeling.

1. Introduction

Designing industrial energy systems given a multi-year planning horizon is challenging yet crucial for achieving a sustainable energy supply [1]. Energy systems are commonly designed to minimize total annualized cost (TAC), the global warming impact, or both. Long planning horizons suffer from a lack of good forecasts and also involve inherent uncertainties, which are recommended to be considered [1,2]. Optimization-based techniques for identifying the optimal energy system configuration under uncertainty include robust optimization, stochastic optimization, or chance-constrained optimization [3,4]. Here, the consideration of uncertainties lead to complex optimization problems that are challenging to solve resulting in a tradeoff between solution quality and computational tractability [3]. The presence of multiple uncertain factors, e.g., arising from the long-term evolution of energy prices and demand, adds further complexity in identifying the best suited design [3]. Energy system optimization under uncertainty can also be addressed using Monte Carlo analysis (MCA), where computation time is independent of the number of uncertain parameters for a single scenario [4,5]. However, MCA typically requires evaluating hundreds of scenarios while the computational time still increases linearly with the number of scenarios and thus can become prohibitively high, particularly when complex models are involved [5]. A promising approach is to use approximation methods involving sampling techniques like Latin hypercube sampling, which can facilitate the consideration of a wide range of scenarios [6,7]. Furthermore, MCA can be quite sensitive to probability distributions for uncertain parameters [5]. Even

https://doi.org/10.52202/069564-0223

though only a few parameters might be influential, the specification of uncertainties is challenging and should be done systematically [2]. Overall, the issue of efficiently incorporating multiple sources of uncertainties into the energy system design process for industrial systems with a multi-year planning horizon is challenging and needs to be addressed.

In this work, we propose a method that helps to make informed decisions regarding the selection of the optimal distributed energy system (DES) design from a set of promising designs considering long time horizons and multiple sources of uncertainty. This so-called Rapid Monte Carlo analysis (RMCA) incorporates uncertainty and surrogate modeling. The economic metric TAC is used to rank the designs. As a first step, we define the procedure of calculating TAC. Determining TAC for a design involves estimating operational expenditures by solving operational optimizations and using investment cost correlations. Accordingly, the metric depends on a variety of uncertain parameters that may influence the results. To model uncertainties, the full set of uncertain parameters is reduced to a few representative factors. Each factor is classified and parameterized to define the probability distributions of the long planning horizon. Next, computationally burdensome relationships between these representative factors and the TAC are replaced by linear surrogate models. The uncertainty and surrogate models allow a large number of long-term scenarios to be considered. Finally, a statistical analysis of the TAC of the DES designs is conducted to serve as a foundation for selecting a final design.

The paper is structured as follows. Section 2 presents the proposed method in detail by first providing an overview, then presenting the uncertainty parameterization and modeling as well as the linear surrogate model generation, and finally the RMCA. Section 3 applies the method to a case study adapted from the literature and discusses the results. Finally, in Section 4 conclusions are drawn.

2. Rapid Monte Carlo analysis via uncertainty and surrogate modeling

2.1. Method overview

We propose a method to analyze a set of promising DES design alternatives to identify the best design given a multi-year planning horizon and multiple sources of uncertainty. To rate a system's performance, we use the total annualized cost (TAC) as assessment metric. The given design alternatives can be, for example, user-defined or the result of an optimization-based design method. Figure 1 shows the four main steps of using our method.

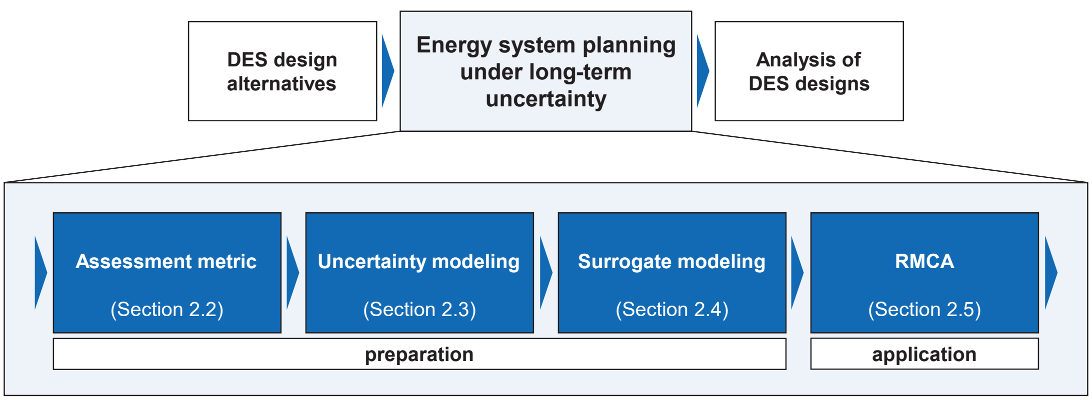

Figure 1. Energy system planning under long-term uncertainty: Steps for preparing and applying the proposed rapid Monte Carlo analysis (RMCA) via uncertainty and surrogate modeling for analyzing promising distributed energy system (DES) design alternatives to identify the best system according to an assessment metric.

Section 2.2 discusses the calculation of the assessment metric TAC and the involved optimization models. TAC depends not only on parameters such as investment costs but also on the operational expenditure (OPEX) including energy costs occurring within the multi-year planning horizon. Here, we consider the OPEX for each year of the planning horizon individually rather than resorting to only one year that is assumed to be representative for the whole horizon. We formulate mixed-integer linear programs (MILPs) for each DES design to determine the energy costs of each year. Each MILP depends on a large number of parameters (e.g., component efficiencies, heating demand for each hour of the year). As a result, the assessment metric depends on many potentially uncertain parameters.

In Section 2.3 we show, how we address uncertainties. We first identify a small number of representative factors that scale reference parameters to define the parameters of the assessment metric. We model the time-dependent development of these representative factors using probability distributions. After modeling uncertainties, probability distributions for all parameters of the TAC are defined. A direct calculation of the TAC

https://doi.org/10.52202/069564-0223

involves solving a MILP for each year of the multi-year planning horizon. Thus, calculating the TAC directly is computationally burdensome. Instead of solving multiple MILPs to obtain the energy costs for each year, we employ linear surrogate models.

Section 2.4 shows how we first generate Latin hypercube samples (LHS) of representative factor combinations to obtain the energy costs of representative years. We then use each combination to parameterize and solve MILPs. Afterwards, we use the LHS to parameterize linear surrogate models. These linear surrogate models allow approximating the energy costs of a single year given the representative factors specifying the respective year. The surrogate model generation is conducted for each DES design. As a result, we can approximate the TAC for long-term scenarios without solving additional MILPs. After following these three preparation steps, the RMCA is carried out.

In Section 2.5, we present the steps of conducting our RMCA. It relies on the probability distributions determined for uncertainty modeling and approximates the TAC using the generated surrogate models for evaluating the long-term scenarios. The RMCA provides decision support by evaluating a large number of multi-year scenarios and conducting a subsequent statistical analysis.

2.2. Assessment metric and energy system modeling

We evaluate DES design alternatives using the TAC as a metric where we consider the OPEX of each year individually, rather than using one representative year. The DES to be evaluated involve different technologies for converting and storing energy. The TAC of an energy system e is calculated as follows:

$$TAC_e = \frac{1}{PVF} * \left(I_{0,e} + \sum_{a=1}^{T} \frac{OPEX_{e,a}}{(1+i)^a} \right) \tag{1}$$

with the net present value factor PVF, the investment costs $I_{0,e}$ of the energy system e, the operational expenditures $OPEX_{e,a}$ of energy system e occurring in year $a \in \{1, \dots, T\}$, the time horizon length of T years, and the interest rate i. We determine the net present value factor as follows [8]:

$$PVF = \frac{(1+i)^T - 1}{(1+i)^T i} \tag{2}$$

We estimate the investment costs $I_{0,e}$ using cost correlations (e.g., cf. [9]). The OPEX $OPEX_{e,a}$ encompasses energy costs $OPEX_{e,a}^{\mathrm{e}}$ as well as maintenance costs $OPEX_{e,a}^{\mathrm{m}}$:

$$OPEX_{e,a} = OPEX_{e,a}^{\mathrm{e}} + OPEX_{e,a}^{\mathrm{m}} \tag{3}$$

The maintenance costs $OPEX_{e,a}^{\mathrm{m}}$ of energy system e are determined in accordance to [9]. To compute the energy costs $OPEX_{e,a}^{\mathrm{e}}$ of energy system e in year a, we utilize MILPs to determine the cost-optimal energy system operation to fulfil the energy demand. The objective function of each MILP is defined as the sum of costs for purchasing energy carriers in year a. Here, we assume that the costs of each year can be computed independently and obtain the energy costs $OPEX_{e,a}^{\mathrm{e}}$ by solving the respective MILP. We model the energy systems using quasi-stationary component models as used by [9] and consider part-load behavior as in [10]. As a result, the MILPs comprise equations for the objective function, energy balances, the conversion units, and the energy storage. The models for the operational optimization of energy systems encompass a variety of parameters (e.g., hourly-resolved heating demands, components' efficiencies). TAC_e of each energy system e can be determined if the parameters are known; however, some parameters like energy prices can be subject to uncertainty.

2.3. Uncertainty parameterization and long-term uncertainty modeling

The assessment metric usually depends on a large number of parameters which are potentially uncertain. We reduce the number of potentially uncertain parameters by first screening and then grouping them. Specifically, we first identify the parameters that can be assumed as constant. Afterward, we group the remaining parameters and assign each group to a representative factor that captures the shared variability. This screening and grouping, to which we refer to as uncertainty parameterization, results in a low number of representative factors. For each representative factor, we then conduct the long-term uncertainty modeling by choosing one of three uncertainty types and defining a permitted range for each representative factor. Thereafter, the uncertainty modeling is complete. The proposed procedure guides the uncertainty modeling steps (i.e., how uncertainty is modeled) and leaves case specific decisions to the user (e.g., which sources of uncertainty should be considered).

We start the uncertainty parameterization with a preliminary screening by first listing all parameters of the assessment metric. Next, parameters with known values, with negligible influence on the assessment metric, or with negligible uncertainty are set as constant. At this point, the risk of excluding influential parameters should be considered carefully and expert knowledge about characterizing uncertainty can be incorporated. For parameters assumed to be uncertain, we determine reference parameters. Following that, the uncertain parameters that are expected to be correlated (e.g., all hourly demands of one year) are grouped and assigned to the representative factor of that group. As soon as all parameters of the assessment metric are either set as constant or assigned to a representative factor, the uncertainty parameterization is complete.

For the long-term uncertainty modeling, we distinguish three types of long-term uncertainty. The three types are adapted from [2]. Figure 2 illustrates the three types of uncertain parameter developments.

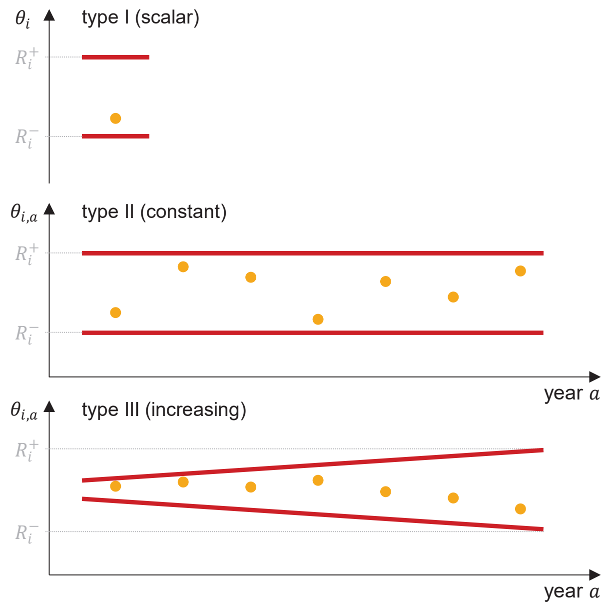

Figure 2. Three types of long-term uncertainty. Type I: Random scalar (e.g., investment costs, life expectancy of equipment). Type II: Constant long-term uncertainty (e.g., energy prices). Type III: Increasing long-term uncertainty (e.g., energy demand). The exemplary realizations (orange dots) of the representative factors (θ_i, $\theta_{i,a}$) lay within the permitted range which is defined by its bounds (R_i^+, R_i^-).

For each representative factor $i \in I = I^{\mathrm{I}} \cup I^{\mathrm{II}} \cup I^{\mathrm{III}}$, first, an appropriate long-term uncertainty type needs to be chosen. Second, a range of permitted scaling factor values needs to be selected. To parameterize the permitted range, an upper bound R_i^+ and a lower bound R_i^- need to be selected. These bounds can, for example, be derived from historical data or forecasts.

We derive the probability distributions of the representative factors as follows. We use normal (Gaussian) distributions, which are one of the most frequently used distributions [11], to model the uncertain parameters. The expected value μ_i of the normal distribution is assumed for all representative factors $i \in I$ as

$$\mu_i = \frac{R_i^+ + R_i^-}{2} \qquad \forall i \in I \tag{4}$$

The standard deviation σ_i is determined such that the permitted range matches the two-sigma interval:

$$\sigma_i = \frac{R_i^+ - R_i^-}{4} \qquad \forall i \in I \tag{5}$$

For type I factors $i \in I^{\mathrm{I}}$, we define the normal random variable $\tilde{\theta}_i$ that serves as an auxiliary variable:

$$\tilde{\theta}_i \sim N(\mu_i, \sigma_i^2) \qquad \forall i \in I^{\mathrm{I}} \tag{6}$$

Analogously, for type II factors $i \in I^{\mathrm{II}}$, we define for each year a an auxiliary variable $\tilde{\theta}_{i,a}$:

$$\tilde{\theta}_{i,a} \sim N(\mu_i, \sigma_i^2) \qquad \forall a \in \{1, \ldots, T\}, i \in I^{\mathrm{II}} \tag{7}$$

We project these auxiliary variables to values within the permitted range. Thus, we obtain the probability distributions for the representative factors $\theta_{i,a}$ as follows:

$$\theta_{i,a} = \begin{cases} R_i^+ & \tilde{\theta}_{i,a} > R_i^+ \\ R_i^- & \tilde{\theta}_{i,a} < R_i^- \\ \tilde{\theta}_{i,a} & \text{otherwise} \end{cases} \qquad \forall a \in \{1, \ldots, T\}, i \in I^{\mathrm{II}} \tag{8}$$

An analogous formulation for $\theta_i \in I^I$ (i.e., one where the index a is dropped) is used for type I parameters. The increasing uncertainty of type III factors $i \in I^{\mathrm{III}}$ is modeled using

$$\tilde{\theta}_{i,a} = \theta_{i,a-1} + \Delta_{i,a} \qquad \forall a \in \{1, \ldots, T\}, i \in I^{\mathrm{III}} \tag{9}$$

 https://doi.org/10.52202/069564-0223

with $\theta_{i,0} = \mu_i \;\forall\, i \in I^{\mathrm{III}}$ and the random variable $\Delta_{i,a} \sim N\left(0, \sigma_{\Delta,i}^2\right) \forall\, a \in \{1,\ldots,T\}, i \in I^{\mathrm{III}}$ which corresponds to the difference of consecutive factor values. We determine the standard deviation $\sigma_{\Delta,i}$ as follows. If random variables are uncorrelated, the variance of their sum equals the sum of their variances [11]. To approximate the variance σ_i^2 for the final parameter $\theta_{i,T}$, we set the variance of each step to $1/T$-th of the final, desired variance. Thus, we determine the respective standard deviations using $\sigma_{\Delta,i} = \sigma_i/\sqrt{T}\;\forall i \in I^{\mathrm{III}}$. After assigning each uncertain factor to an uncertainty type and parameterizing it, long-term scenarios can be generated, and the respective TAC can be computed.

2.4. Surrogate modeling via Latin hypercube sampling and linear regression

We use linear surrogate models to speed up the calculation of the assessment metric by replacing computational burdensome operations with linear approximations. For determining the TAC of a DES design, the energy costs need to be calculated for each year of the multi-year time horizon of length T. As a result, T optimization problems need to be solved for each long-term scenario. After introducing the representative factors in Section 2.3, which includes setting parameters to constant values and determining reference parameters, each optimization problem is fully defined by a small set of parameters:

$$OPEX_{e,a}^{\mathrm{e}} = OPEX_{e}^{\mathrm{e}}\left(\{\theta_i\}_{i\in I^{\mathrm{I}}}, \{\theta_{i,a}\}_{i\in I^{\mathrm{II}}\cup I^{\mathrm{III}}}\right) \tag{10}$$

For each DES design, we obtain a linear surrogate model by first solving the respective MILP for a representative set of sampling points. We select these sampling points using Latin hypercube sampling and solve the MILPs for N^{LHS} combinations of representative factors. Afterwards, we use ordinary least squares [12] to parameterize the linear surrogate models which have the following form:

$$OPEX_{e,a}^{\mathrm{e}} \approx \widetilde{OPEX}_{e,a}^{\mathrm{e}} = b_{e,0} + \sum_{i\in I^{\mathrm{I}}} b_{e,i}\theta_i + \sum_{i\in I^{\mathrm{II}}\cup I^{\mathrm{III}}} b_{e,i}\theta_{i,a} \tag{11}$$

where $\widetilde{OPEX}_{e,a}^{\mathrm{e}}$ is the linear approximation of the energy costs $OPEX_{e,a}^{\mathrm{e}}$ of DES design e in year a and $b_{e,i}$ is a surrogate model parameter. We assess the quality of fit using the coefficient of determination R^2 defined as [13]:

$$R^2 = 1 - \frac{\sum_{j=1}^{N^{\mathrm{LHS}}}\left(X_j - Y_j\right)^2}{\sum_{j=1}^{N^{\mathrm{LHS}}}\left(\bar{Y} - Y_j\right)^2} \quad \text{with} \quad \bar{Y} = \frac{\sum_{j=1}^{N^{\mathrm{LHS}}} Y_j}{N^{\mathrm{LHS}}} \tag{12}$$

where Y_j is the value of the observed data (i.e., the result of an optimization), X_j is the value predicted by a linear surrogate model, $\bar{Y}$ is the mean of the observed data, and N^{LHS} corresponds to the number of Latin hypercube samples. The coefficient of determination can take values $R^2 \leq 1$ and positive values of R^2 can be linked to the percentage of correctness of a regression [13]. However, a high value of R^2 cannot guarantee whether the linear surrogate model is an appropriate approximation. After having generated linear surrogate models, we approximate the TAC as follows:

$$TAC_e \approx \widetilde{TAC}_e = \frac{1}{PVF} * \left(I_{0,e} + \sum_{a=1}^{T} \frac{\widetilde{OPEX}_{e,a}^{\mathrm{e}} + OPEX_{e,a}^{\mathrm{m}}}{(1+i)^a}\right) \tag{13}$$

2.5. Rapid Monte Carlo analysis

Our rapid Monte Carlo analysis (RMCA) allows to analyze and compare promising DES design alternatives with a reduced computation burden considering multiple sources of uncertainty and long planning horizons. The RMCA is enabled by defining an assessment metric (Section 2.2), conducting uncertainty parameterization as well as long-term uncertainty modeling (Section 2.3), and generating surrogate models (Section 2.4). Our RMCA involves the typical steps of a MCA [5]. First, probability distributions are defined. Second, random values, i.e., long-term scenarios, are sampled. We use long-term uncertainty models to generate multi-year scenarios defined by a set of representative factors. Third, the samples are evaluated using approximations for the TAC. By using linear surrogate models, the computational burden of approximating the TAC can be neglected. After generating and evaluating a sufficiently large number of long-term scenarios, we conduct a statistical analysis to identify the best DES design alternative.

3. Case study and results

3.1. Distributed energy system design alternatives for a research campus

We apply our method in a case study adopted from the literature. In our case study, we seek to find the best DES design from a set of promising candidates to fulfill energy demands, i.e., the heating and electricity demands given in [9].

Figure 3 shows three user-defined DES design alternatives (i.e., ES1, ES2, ES3) which are to be analyzed. The designs are similar to the those obtained using an energy system design optimization in [9]. All designs include two boilers and a thermal energy storage. ES1 includes a combined heat and power (CHP) unit and a heat pump (HP), ES2 has two CHP units but no HP, and ES3 has two HPs instead.

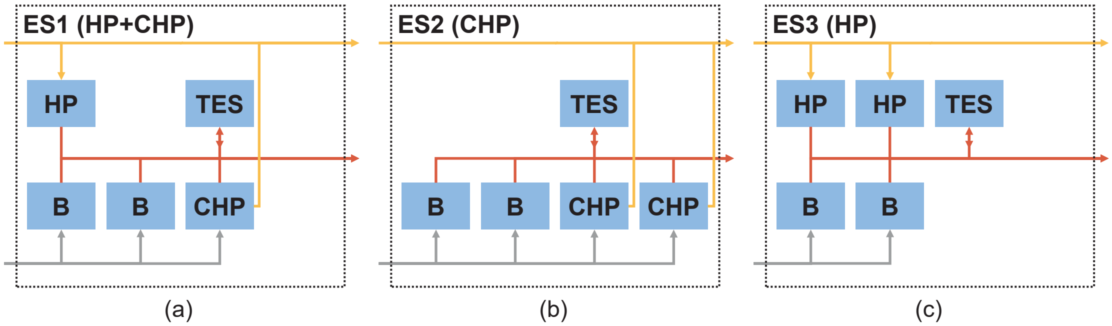

Figure 3. Promising distributed energy system design alternatives $e \in \{ES1, ES2, ES3\}$. The three designs encompass boilers (B), heat pumps (HP), combined heat and power (CHP) units, and thermal energy storages (TES) to fulfill an electricity as well as a heating demand. The systems are connected to an electricity and a natural gas grid.

Technical specifications of the conversion units are shown in Table 1. The part-load segment parameters ($\lambda_{k,\min}^{\text{in/out}}$; $\lambda_{k,\max}^{\text{in/out}} = 1$) are derived using correlations given in [9]. The thermal energy storage has for all cases a capacity of $Q^{\max} = 1000\text{ kWh}$, charging efficiencies of $\eta^{\text{in}} = \eta^{\text{out}} = 0.95$, and a heat loss time constant of $\tau^{\text{loss}} = 200\text{ h}$.

Table 1. Nominal power, nominal efficiency, and part-load segment parameters of the installed boilers (B), combined heat and power (CHP) units, and heat pumps (HP).

	Nominal power	Nominal efficiency	$\lambda_{k,\min}^{\text{in}}$	$\lambda_{k,\min}^{\text{out}}$
Boiler	$\dot{Q}_{\text{B}}^{\text{th}} = 530\text{ kW}$	$\eta_{\text{B}}^{\text{th}} = 80.0\%$	0.173	0.200
CHP unit (electric)	$P_{\text{CHP}}^{\text{el}} = 380\text{ kW}$	$\eta_{\text{CHP}}^{\text{el}} = 38.9\%$	0.582	0.500
CHP unit (thermal)	$\dot{Q}_{\text{CHP}}^{\text{th}} = 470\text{ kW}$	$\eta_{\text{CHP}}^{\text{th}} = 48.1\%$	0.582	0.622
Heat pump	$\dot{Q}_{\text{HP}}^{\text{th}} = 200\text{ kW}$	$COP = 2.998$	0.200	0.200

The investment costs and annual maintenance costs of each DES design alternative are determined using the cost correlations given in [9] and are shown in Table 2.

Table 2. Investment costs and annual maintenance costs of each DES design alternative.

	ES1 (HP+CHP)	ES2 (CHP)	ES3 (HP)
Investment costs $I_{0,e}$, €	383,901	551,542	216,260
Maintenance costs $OPEX_{e,a}^{\text{m}}$, €	24,331	46,089	2,574

Given the DES design alternatives, we conduct the first of the three preparation steps shown in Fig. 1: We define for each DES design an assessment metric according to Eq. (1) for a planning horizon length of $T = 10$ years. This step includes the formulation of MILPs to determine annual energy costs.

3.2. Uncertainty and surrogate modeling

Next, we present the preparation of the uncertainty and surrogate models for each DES design alternative. We start modeling uncertainty by listing all parameters influencing the assessment metric. As the designs include different components, the assessment metrics of the design alternatives depend on different sets of model parameters. Next, we identify the uncertain parameters. As the price of energy carriers can have a relevant impact in energy planning models [2], we model the gas price and the electricity price as time-dependent, uncertain parameters. Furthermore, we assume that also the heating demand and the electricity demand are relevant uncertain parameters. All other model parameters are set to appropriate constant values. Hence, due to the hourly resolution of the optimization models, a total number of $4 \cdot 8760$ parameters per year remain uncertain. We introduce three representative factors, i.e., one factor for scaling electricity prices ($\theta_{\text{c-el},a}$), one factor for scaling gas prices ($\theta_{\text{c-gas},a}$), and one factor for scaling electricity and heating demands ($\theta_{\text{d},a}$). We assign each of the $4 \cdot 8760$ uncertain parameters to one representative factor for scaling reference parameters (cf. Table 3). The representative factors serve as a low-dimensional representation of the uncertain parameters reducing the number of uncertain parameters to three per year.

https://doi.org/10.52202/069564-0223

Table 3. Representative factors and long-term uncertainty modeling. For all representative factors, the associated uncertain parameters of the assessment metric, the type of uncertainty, and the bounds (R_i^+, R_i^-) of the permitted range are given.

Representative factor	Associated parameters	Type of uncertainty	R_i^-	R_i^+
$\theta_{\mathrm{c-el},a}$	Hourly resolved electricity price	II (constant)	0.44	1.57
$\theta_{\mathrm{c-gas},a}$	Hourly resolved gas price	II (constant)	0.26	1.74
$\theta_{\mathrm{d},a}$	Hourly resolved electricity and heat demand	III (increasing)	0.60	1.40

We classify the representative factor $\theta_{\mathrm{d},a}$ as type III uncertainty as demand uncertainties typically increase over time. Furthermore, we assume that the demand might increase or decrease by up to 40 % over the time horizon of 10 years. This assumption results in an upper bound of $R_{\mathrm{d}}^+ = 1.4$ and a lower bound of $R_{\mathrm{d}}^- = 0.6$. We use the heating and electricity demand from [9] for defining the corresponding reference parameters. As done in [2], we assume that the accuracy of short and long-term energy price predictions is equal. Thus, we classify $\theta_{\mathrm{c-el},a}$ and $\theta_{\mathrm{c-gas},a}$ as type II uncertainties. For modeling the long-term uncertainty of the energy prices, we assume a low-price and a high-price scenario considering historical prices (cf. [14]). In the low-price scenario, we assume an average electricity price of $7.5\ \mathrm{ct/kWh}$ and a gas price of $2.5\ \mathrm{ct/kWh}$. In the high-price scenario, we assume an electricity price of $26.67\ \mathrm{ct/kWh}$ and a gas price of $17.44\ \mathrm{ct/kWh}$. We assume that within one year the energy prices remain constant. We choose the upper bound $R_{\mathrm{c-el}}^+$ such that the reference price parameters scaled with $R_{\mathrm{c-el}}^+$ result in the high-price scenario. We determine the lower bound $R_{\mathrm{c-el}}^-$ and the range of the representative factor $\theta_{\mathrm{c-gas},a}$ analogously. After defining a low-dimensional parameterization and modeling long-term uncertainty, we generate the surrogate models.

The linear surrogate models for approximating the energy costs of one year of operation are generated by first solving the respective MILPs for $N^{\mathrm{LHS}} = 25$ Latin hypercube samples and then applying linear regression to obtain the parameters of the linear surrogate models. We use Pyomo 6.4 [15,16] for modeling and Gurobi 10.0 [17] as a solver. Table 4 shows the linear surrogate model parameters for each of the DES design as well as the quality of fit given by the R^2-value.

Table 4. Linear surrogate models for approximating the annual energy costs of the energy systems. The linear surrogate models are defined by its parameters $b_{e,i}$. The coefficient of determination R^2 aids assessing the model quality.

e	$b_{e,0}$, €	$b_{e,c\mathrm{-el}}$, €	$b_{e,c\mathrm{-gas}}$, €	$b_{e,\mathrm{d}}$, €	R^2
ES1	-520,442	239,815	219,318	537,369	0.9429
ES2	-475,527	224,834	296,459	524,416	0.9617
ES3	-556,080	396,682	56,741	594,618	0.9746

Figure 4 shows by comparing observed and approximated energy costs that energy costs are predicted reasonably well by the linear surrogate models for each LHS and DES design.

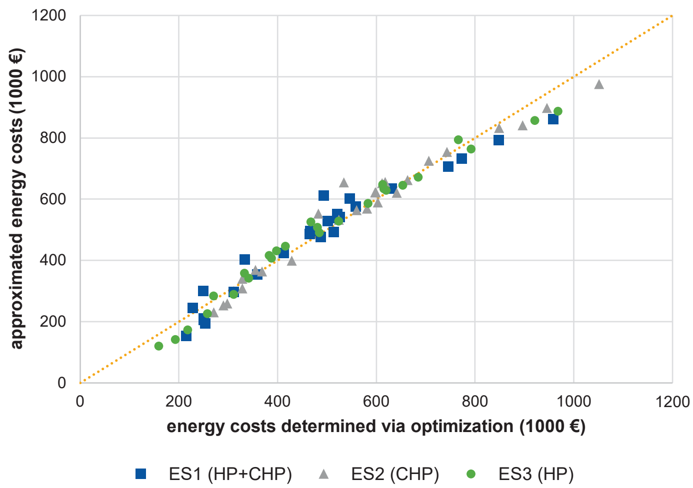

Figure 4. Scatter plot for assessing the accuracy of the three linear surrogate models based on the Latin hypercube samples. Small vertical deviations from the dotted line indicate a high approximation accuracy.

3.3. Rapid Monte Carlo analysis

The RMCA is applied employing the uncertainty and surrogate models. Our analysis is presented in three parts. First, we present the generated scenarios, which consist of $N = 5000$ long-term scenarios illustrated in Fig. 5. Next, we provide a statistical analysis from three views to evaluate the scenarios. Finally, we consider the computational burden associated with the RMCA.

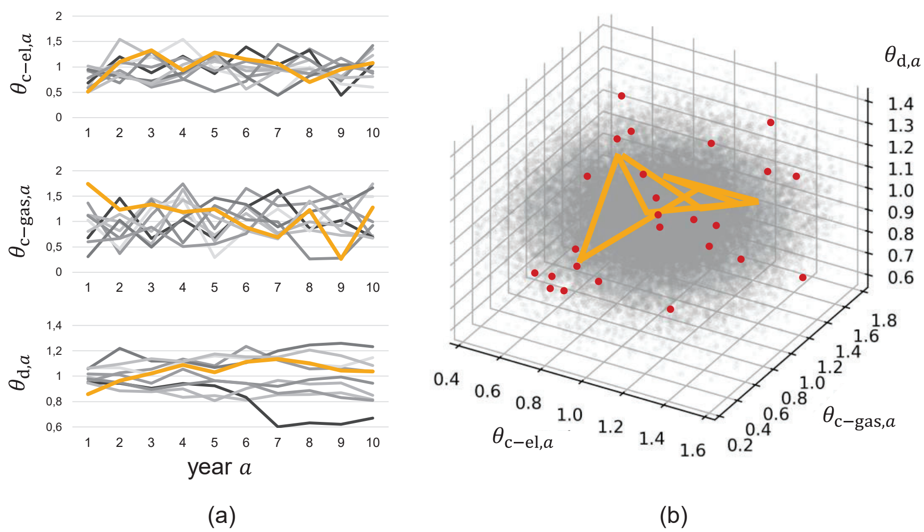

Figure 5. Long-term scenarios ($N = 5000$) and Latin hypercube samples ($N^{\mathrm{LHS}} = 25$). a) The subset of ten long-term scenarios showing different trajectories of the representative factors ($\theta_{\mathrm{c-el},a}$, $\theta_{\mathrm{c-gas},a}$, $\theta_{\mathrm{d},a}$). b) The parameter space is spanned by the representative factors and is densely covered by the 50,000 years contained within the long-term scenarios (small gray dots). One long-term scenario and the Latin hypercube samples are illustrated in orange and by the red dots, respectively.

We evaluate the long-term scenarios using the linear surrogate models and thus avoid solving many optimization problems with similar parameterizations. We analyze the TAC obtained for the long-term scenarios and DES design alternatives from three views. First, we analyze the distributions of the TAC for each design.

https://doi.org/10.52202/069564-0223

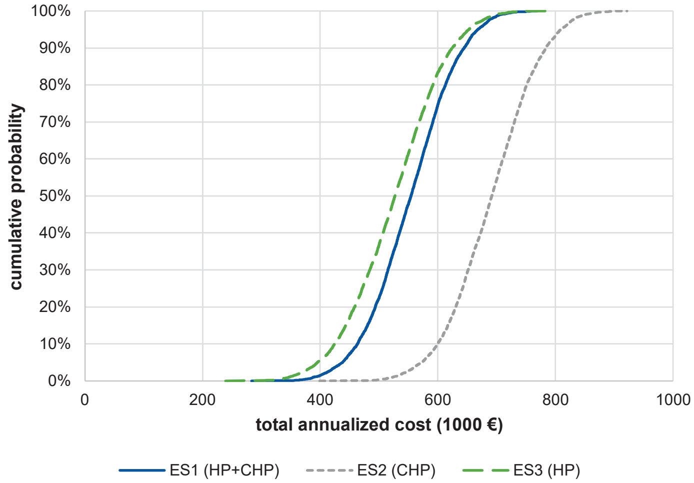

Figure 6. Cumulative probability plot for determining the probability of the total annualized cost (TAC) being lower or equal to the given TAC when choosing the respective design.

Figure 6 reveals the bandwidth of expected TAC. This view indicates that ES2 is likely to result in the highest TAC and that furthermore ES1 and ES3 lead to similar TAC. However, this visualization does not include information about the difference in TAC for specific long-term scenarios.

By comparing the TAC of the DES design alternatives in each scenario, we determine the frequency of having the lowest TAC for each design. We find that ES3 results in 87.7 % of the scenarios in the lowest TAC. In the remaining scenarios, ES1 results in the lowest TAC. However, this second view does not convey information about the differences in TAC, i.e., how high the regret is when another design would have been a better choice.

Third, we examine the differences in TAC. We normalize the differences using the expected TAC of ES3. We find that the maximal regret associated with choosing ES1, ES2, and ES3 is 19.4 %, 51.4 %, and 11.4 %, respectively. In this context, regret is defined as the difference between the TAC of the selected design and the TAC of the best design. Furthermore, we find that the average regret associated with choosing ES1, ES2, and ES3 is 5.5 %, 31.7 %, and 0.3 %, respectively. Thus, we recommend ES3 as it most likely results in the lowest TAC whereas choosing ES2 should be avoided. Overall, the different views of the statistical analysis provide insights about the probability distributions of the assessment metric. Furthermore, analyzing the differences in TAC can reveal which design should be preferred or avoided.

The computational performance of the RMCA is addressed next. Here, we use the number of MILPs to be solved as a proxy for the computational burden. We compare the RMCA to a MCA that does not employ surrogate models. When not using surrogate models, we need to solve one MILP for obtaining the energy costs for each year, scenario, and design. As a result, the computational burden increases linearly with the number of long-term scenarios. Given three DES design alternatives and a time horizon of ten years, each scenario requires solving 30 MILPs. Solving the MILPs of this work typically requires a few minutes. As MCA applied for energy system design might require hundreds of scenarios to be considered [4,5], not using surrogate models results in a significant computational burden.

Using the RMCA also requires solving MILPs. Before we can evaluate the first scenario using RMCA, we need to create the surrogate models first. We solve 75 MILPs for the three linear surrogate models in total. In comparison to solving MILPs, the computational burden of determining the parameters of the linear surrogate models and evaluating them is negligible. As a result, the RMCA is faster if three or more long-term scenarios are considered. For the considered 5,000 scenarios, the computational time is reduced by more than 99 %.

4. Conclusion

Designing distributed energy systems (DES) under uncertainty is a challenging task. One approach to address uncertainty is to use Monte Carlo analysis (MCA). However, the high computational burden associated with MCA for DES design and long planning horizons limits its application. Furthermore, MCA requires reasonable assumptions about probability distributions. To address the challenge of designing DES under uncertainty considering multi-year planning horizons, we propose a rapid Monte Carlo analysis (RMCA) approach enabled by long-term uncertainty modeling and surrogate modeling for optimization-based energy system planning under long-term uncertainty. Our method provides decision support when a set of promising DES designs is given.

To enable RMCA, three preparation steps are conducted. Firstly, an assessment metric is defined and mixed-integer linear programs (MILPs) for the operational optimization of each DES design are formulated. Secondly, a low-dimensional representation of uncertainty and a long-term uncertainty model are systematically derived. Thirdly, surrogate models are created after solving the respective MILP for representative Latin hypercube samples. The RMCA starts by generating a large number of scenarios by sampling the long-term uncertainty models. The scenarios are evaluated using the surrogate models. Subsequently, a statistical analysis reveals the expected distribution of the total annualized cost (TAC) for each design, the likelihood of a specific design resulting in the lowest TAC, and the regret of choosing one design over another.

We apply our method to a case study adapted from the literature to assess three promising DES designs for a time horizon of ten years. We formulate MILPs for the operational optimization of each DES design and represent uncertain parameters using three representative factors per year. We create accurate linear surrogate models with a coefficient of determination $R^2 \geq 0.94$ for all designs. In the RMCA, 5,000 scenarios are considered and the best design which most likely results in the lowest costs is identified. The RMCA approach is compared to a MCA that does not rely on surrogate models and thus does not require solving MILPs for representative samples but lacks the benefit of the low effort of evaluating a surrogate model. Using the surrogate models results in a speedup if three or more long-term scenarios are considered. When considering 5,000 scenarios, the computational time is reduced for the specific case study by more than 99 %.

The presented approach can be applied to a wide range of energy systems. Besides, created surrogate models could also be used in companioning sensitivity analyses, which would also benefit from the substantially reduced computational burden. The benefit of a sensitivity analysis is that it does not require information about probability distributions. For DES designs of higher complexity, nonlinear surrogate models might be required to reach a high quality of fit, increasing the effort of creating the surrogate models.

In summary, the RMCA approach enabled by uncertainty and surrogate modeling provides an efficient way of considering multiple sources of uncertainty and long planning horizons. The adaptability and the reusability of its components, i.e., of the energy system, the uncertainty, and the surrogate modeling, potentially enable wide applicability of the proposed approach for energy system design under uncertainty.

Acknowledgments

This study is funded by the German Federal Ministry of Economic Affairs and Energy (ref. no.: 03EN2031A). The support is gratefully acknowledged.

Nomenclature

COP	coefficient of performance, -
i	interest rate, -
$I_{0,e}$	investment costs of energy system e, €
N	number of long-term scenarios, -
N^{LHS}	number of Latin hypercube samples, -
$OPEX_{e,a}$	operational expenditure of energy system e in year a, €
$OPEX_{e,a}^{\mathrm{m}}$	maintenance costs of energy system e in year a, €
$OPEX_{e,a}^{\mathrm{e}}$ / $\widetilde{OPEX}_{e,a}^{\mathrm{e}}$	energy costs of energy system e in year a (observed/approximated), €
PVF	net present value factor, -
$\dot{Q}_k^{\mathrm{th}}$ / P_k^{el}	nominal thermal/electrical capacity of unit k, kW
$Q^{\max}$	capacity of thermal storage, kWh
R^2	coefficient of determination, -
R_i^+ / R_i^-	upper/lower bound of the permitted range of representative factor i, -
T	time horizon length, -
TAC_e / $\widetilde{TAC}_e$	total annualized cost of energy system e (observed/approximated), €
η_k^{th} / η_k^{el}	nominal thermal/electrical efficiency of unit k, -
$\eta^{\mathrm{in/out}}$ / τ^{loss}	efficiency/heat loss time constant of thermal storage, -
θ_i / $\theta_{i,a}$	representative factor i in year a, -
$\tilde{\theta}_i$ / $\tilde{\theta}_{i,a}$ / $\Delta_{i,a}$	auxiliary variable for representative factor i in year a, -
$\lambda_{k,\mathrm{min/max}}^{\mathrm{in/out}}$	part-load efficiency parameter of unit k, -
μ_i	expected value of auxiliary variable for representative factor i, -
σ_i / $\sigma_{\Delta,i}$	standard deviation of auxiliary variable for representative factor i, -

https://doi.org/10.52202/069564-0223

References

[1] Mavromatidis G, Petkov I. MANGO: A novel optimization model for the long-term, multi-stage planning of decentralized multi-energy systems. Applied Energy 2021;288:116585. https://doi.org/10.1016/j.apenergy.2021.116585.

[2] Moret S, Codina Gironès V, Bierlaire M, Maréchal F. Characterization of input uncertainties in strategic energy planning models. Applied Energy 2017;202:597–617. https://doi.org/10.1016/j.apenergy.2017.05.106.

[3] Roald LA, Pozo D, Papavasiliou A, Molzahn DK, Kazempour J, Conejo A. Power systems optimization under uncertainty: A review of methods and applications. Electric Power Systems Research 2023;214:108725. https://doi.org/10.1016/j.epsr.2022.108725.

[4] Fodstad M, Crespo del Granado P, Hellemo L, Knudsen BR, Pisciella P, Silvast A, et al. Next frontiers in energy system modelling: A review on challenges and the state of the art. Renewable and Sustainable Energy Reviews 2022;160:112246. https://doi.org/10.1016/j.rser.2022.112246.

[5] Yue X, Pye S, DeCarolis J, Li FGN, Rogan F, Gallachóir BÓ. A review of approaches to uncertainty assessment in energy system optimization models. Energy Strategy Reviews 2018;21:204–17. https://doi.org/10.1016/j.esr.2018.06.003.

[6] Nolting L, Spiegel T, Reich M, Adam M, Praktiknjo A. Can energy system modeling benefit from artificial neural networks? Application of two-stage metamodels to reduce computation of security of supply assessments. Computers & Industrial Engineering 2020;142:106334. https://doi.org/10.1016/j.cie.2020.106334.

[7] Köhnen CS, Priesmann J, Nolting L, Kotzur L, Robinius M, Praktiknjo A. The potential of deep learning to reduce complexity in energy system modeling. Intl J of Energy Research 2022;46:4550–71. https://doi.org/10.1002/er.7448.

[8] Ross SA, Westerfield R, Jaffe JF. Corporate finance. Eleventh Edition. New York, NY: Published by McGraw-Hill Education; 2016.

[9] Sass S, Faulwasser T, Hollermann DE, Kappatou CD, Sauer D, Schütz T, et al. Model compendium, data, and optimization benchmarks for sector-coupled energy systems. Computers & Chemical Engineering 2020;135:106760. https://doi.org/10.1016/j.compchemeng.2020.106760.

[10] Voll P, Klaffke C, Hennen M, Bardow A. Automated superstructure-based synthesis and optimization of distributed energy supply systems. Energy 2013;50:374–88. https://doi.org/10.1016/j.energy.2012.10.045.

[11] Papoulis A, Pillai SU. Probability, random variables, and stochastic processes. 4th ed. Boston: McGraw-Hill; 2002.

[12] Seabold S, Perktold J. statsmodels: Econometric and statistical modeling with python. 9th Python in Science Conference 2010.

[13] Chicco D, Warrens MJ, Jurman G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Computer Science 2021;7:e623. https://doi.org/10.7717/peerj-cs.623.

[14] BDEW. Strompreisanalyse, Februar 2023. Haushalte und Industrie 2023. https://www.bdew.de/media/documents/230215_BDEW-Strompreisanalyse_Februar_2023_15.02.2023.pdf (accessed February 24, 2023).

[15] Bynum ML, Hackebeil GA, Hart WE, Laird CD, Nicholson BL, Siirola JD, et al. Pyomo — Optimization Modeling in Python. vol. 67. Cham: Springer International Publishing; 2021. https://doi.org/10.1007/978-3-030-68928-5.

[16] Hart WE, Watson J-P, Woodruff DL. Pyomo: modeling and solving mathematical programs in Python. Math Prog Comp 2011;3:219–60. https://doi.org/10.1007/s12532-011-0026-8.

[17] Gurobi Optimization, LLC. Gurobi Optimizer Reference Manual 2023.