A Metamodeling Framework for Accelerated Energy Market Optimization Using Active Learning

Benjamin Uhrich[1,2](✉), Linus Thrän[3], and Felix Böing[3]

[1] Leipzig University, Leipzig, Germany
[2] Center for Scalable Data Analytics and Artificial Intelligence Dresden/Leipzig, Leipzig, Germany
uhrich@informatik.uni-leipzig.de
[3] 50Hertz Transmission GmbH, Berlin, Germany

Abstract. The increasing transformation of the European energy market, driven by the rise of intermittent renewable energies, the decommissioning of controllable power plants and dependence on short-term storage, poses challenges for assessing security of supply. In order to evaluate the capacity of available generation to meet demand in uncertain conditions, market model optimizations using the Monte Carlo (MC) approach are employed. However, the high computational costs of this approach limit assessment resolution. This paper investigates metamodeling as a strategy to reduce these computational costs. Metamodelling is a process of using mathematical models on a subset of simulations to map outcomes to the input data. This reduces the total number of simulations required. The study explores three key steps: enhancing input-output correlation, identifying effective machine learning (ML) models and selecting optimal training samples. While no single model performs adequately due to data complexity, a two-model pipeline significantly improves prediction accuracy. An active learning approach is also introduced to further optimize sample selection. The results show that training on only twenty percent of the data reduces the computation time by more than 75%, with a relative error below 10%. These findings demonstrate the potential of metamodeling to enable efficient, high-resolution resource adequacy assessments in evolving energy systems.

Keywords: Metamodeling · energy systems · active learning · resource adequacy · electricity market modeling

1 Introduction

The development of the European energy market presents a fundamental challenge.

In recent years, policy and market developments have prioritized environmental sustainability, sometimes at the expense of economic efficiency and supply

L. Martínez-Villaseñor et al. (Eds.): MICAI 2025, LNAI 16265, pp. 211–227, 2026.
https://doi.org/10.1007/978-3-032-17933-3_22

reliability [1]. However, shifting geopolitical dynamics have redirected public and political attention toward ensuring a robust security of supply [2]. In Germany, the phase-out of all nuclear power plants by 2023 significantly reduced the market's dispatchable capacity. The planned coal phase-out by 2038 will further diminish the available firm capacity.

This reduction has been made possible by the increasing share of renewable energy sources, resulting in a more environmentally friendly energy mix. However, this transition raises new concerns about supply reliability. Ensuring an adequate level of supply, also known as resource adequacy, requires continuous, high-resolution assessments to support well-informed decisions regarding energy market development and design.

Resource adequacy assessments evaluate the frequency with which a given generation system, under the prevailing market design, is capable of meeting electricity demand. By comparing the resulting loss of load expectation (LOLE) with the nationally defined reliability standard, it is possible to determine whether the system is over- or under-dimensioned with respect to security of supply.

Methodologically, it is standard practice to account for uncertainty in both weather conditions and generation availability. This is typically achieved by using multiple historical weather years to reflect variability in renewable generation, and by applying so-called outage draws to simulate the stochastic availability of conventional power plants.

To address these challenges, the agency for the cooperation of energy regulators (ACER) has developed a methodology based on MC simulations run by energy market models. While effective, this approach is computationally intensive, which constrains the resolution and scope of the assessment due to model simplification and data aggregation. For example, the European resource adequacy assessment (ERAA) carries out 700 full-year energy market simulations to evaluate supply reliability [3]. Similar analyses may require approximately six days of computation, motivating the search for alternative or more efficient methodologies.

The present paper explores the use of metamodelling to reduce computational effort. Metamodelling involves the use of mathematical models, frequently derived from ML, to approximate the relationship between input and output data from MC energy market simulations. The training of a metamodel on a subset of simulated data facilitates the estimation of results for unsimulated scenarios, thereby reducing the number of full simulations required.

Despite the evidence that this technique has shown promise in predicting prices from energy market models, its application to capacity sufficiency metrics remains largely unexplored [4]. The issue pertains to the inherent characteristics of the output data. Power systems are usually designed to satisfy demand in most states. Consequently, circumstances where supply is inadequate are rare and difficult to predict in the absence of exhaustive simulation. The identification of such events without the necessity of simulating every possible scenario represents a significant challenge.

This challenge is further compounded by the rise of short-term energy storage, which introduces temporal dependencies to an already spatially interconnected European grid. Furthermore, a robust metamodel must account for the weather dependence of renewable energy sources and incorporate probabilistic modelling of power plant outages.

– Our analysis identifies a high-performing two-model pipeline that minimizes mean absolute error (MAE) in predicting LOLE across MC years.
– We develop an active learning-based approach that iteratively selects the most informative simulation samples to optimize model training efficiency.
– We benchmark the performance of the proposed metamodeling approach against traditional energy market model optimization methods, demonstrating substantial reductions in computational cost (in hours) while maintaining high predictive accuracy (measured by MAE).

2 Related Work

2.1 ML in Energy System Research

ML is increasingly applied in energy system research, though its use in resource adequacy assessments remains relatively unexplored. Starke *et al.* systematically reviewed the role of ML across energy domains, describing it as an emerging yet rapidly growing field [5].

ML has become indispensable in forecasting applications, including load profiles, renewable generation, and energy prices. Rahman *et al.* demonstrated the predictive power of ML in these contexts [6]. Donti *et al.* also proposed ML integration for accelerating and improving energy system optimization tasks [7].

2.2 ML in Energy Model Optimization

ML applications in energy model optimization generally fall into two categories: power flow simulations and economic dispatch. Duchesne *et al.*, Sun *et al.*, and Su *et al.* employed ML to accelerate power flow calculations [8–10]. Leonori *et al.* applied ML in near real-time power flow for microgrids. Huang *et al.* used ML to improve reliability in congested systems [11,12].

In economic dispatch, Darville *et al.* reduced dimensionality in mixed-integer programs to accelerate solver times [13]. Bogensperger *et al.* and Hoffmann *et al.* used ML for time series aggregation to streamline simulations [14,15]. Perera *et al.* developed metamodels to address computational bottlenecks in energy system optimization [16].

2.3 Metamodels in Energy Market Model Optimization

Metamodeling has emerged as a promising alternative to traditional simplification techniques in energy market model optimization. Kohnen *et al.* employed deep learning to map input-output relations and predict market prices [4]. Prina *et al.*, Danish *et al.*, and Kim *et al.* confirmed that metamodels can produce accurate outputs in significantly shorter runtimes, especially under hardware or time constraints [17–19].

Nolting *et al.* showed that ML and design of experiments (DOE) techniques are promising for approximating model behavior with limited computing resources [20]. Priesmann *et al.* supported input data reduction and structural simplification as alternative strategies [21].

2.4 Metamodels in Resource Adequacy Assessments

Resource adequacy assessments remain challenging due to their probabilistic nature, requiring evaluation across diverse weather and outage scenarios. Nolting and Praktiknjo described this as the "complexity dilemma," where increasingly sophisticated questions necessitate complex models. The prediction accuracy of these model depends on the quality of the input data. [22].

By combining the availability curve with residual load, the loss of load probability (LOLP) and LOLE can be estimated, though the process remains computationally expensive [23].

Munch *et al.* advanced this by predicting the availability curve outcomes directly, effectively mitigating the complexity dilemma [24].

Conventional availability curve methods are predicated on situations in isolation, thereby facilitating data generation via DOE. However, contemporary challenges necessitate temporal interconnection modelling due to escalating storage capacity and cross-border interactions. Energy market models offer this, but lack built-in probabilistic capabilities, making MC simulation integration essential but computationally costly.

3 Methodology

3.1 Metamodel Topology

The goal of the metamodel is to predict a LOLE for each market area (MA) with only a fraction of the simulated MC-years having results available. Instead of the standard LOLE (hours with actual shortfall), we define LOLE via the expected number of maximal price events (MPEs).

It eliminates the complexity arising from the expected energy not supplied temporal [25] and spatial [26] distribution effects, which are challenging to predict. Consequently, these effects do not unnecessarily complicate the formation of the metamodel. However, it is important to note that future developments of the approach presented here may address this issue. The metamodel is configured as a pipeline that contains three main parts:

- a classification model,
- an aggregation step, and
- a scaling model.

The classification model's function is to predict whether a MPE occurs for a given situation, as identified by the combination of outage sample (OS), climatic year (CY), MA, and hour. In the aggregation step, the results of the classification model are aggregated as a sum over the hours to obtain a prediction for each MC-year region (identified by the combination of OS, CY, and MA). In a concluding step, the aggregated classification results are scaled. Given that the objective of the scaling model is to only capture systematic over- or underestimation, the implementation of a linear regression model is well-suited for this purpose. The metamodel pipeline can be seen in Fig. 1:

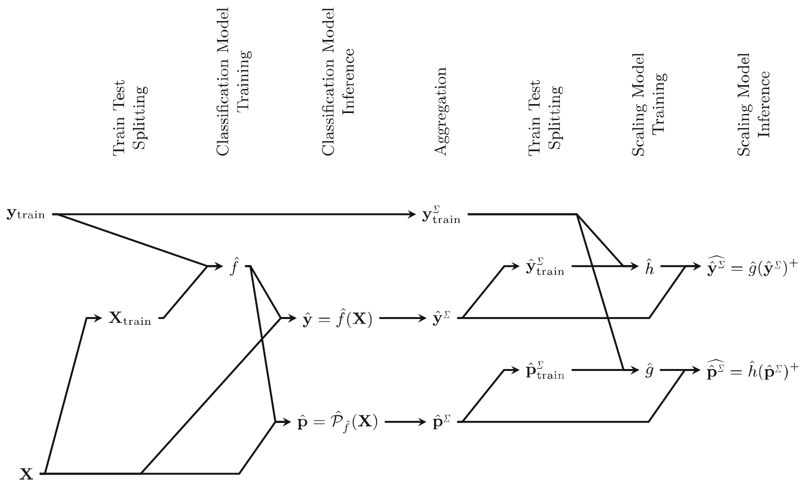

Fig. 1. Metamodel as a pipeline for predicting the LOLE of MC-year regions.

Obvious energy system features such as renewable generation, available conventional capacity or electricity demand are applied, as are newly introduced metrics that take temporal and spatial couplings into account and focus on the remaining capacity margin (available thermal capacity minus residual load). Temporal features are weekend or day of the week as well as hour of the day. In total, we identified 49 features. The input data for the pipeline consists of a single value for each feature in every situation. The input data consists of 700 MC-years in hourly resolution for one scenario and is represented by the matrix $\mathbf{X} \in \mathbb{R}^{n \times F}$. The matrix $\mathbf{X}$ has F columns (one for each feature) and n rows (one for each situation) with $n = \mathbb{H} \cdot |\mathbb{M}| \cdot |MC| = 8760 \cdot 11 \cdot 700$. A subset of the MC-years is selected as the training set $MC_{\text{train}} = MC_i \subset MC$, $|MC_{\text{train}}| = 140$. These MC-years are simulated using the energy market simulation yielding a result vector $\mathbf{y}_{\text{train}} \in \mathbb{R}^{m \times 1}$ with $m = \mathbb{H} \cdot |\mathbb{M}| \cdot |MC_{\text{train}}| = 8760 \cdot 11 \cdot 140$ corresponding to the input data represented by $\mathbf{X}_{\text{train}} \in \mathbb{R}^{m \times F}$.

Classification. The classification model can be fit to approximate the results $\mathbf{y}_{\text{train}}$ as the target regarding the training input data $\hat{f} \approx f : \mathbf{X}_{\text{train}} \to \mathbf{y}_{\text{train}}$. The fit classification model can subsequently be used to predict the target $\hat{\mathbf{y}} = \hat{f}(\mathbf{X})$, or its model internal probability $\hat{\mathbf{p}} = \hat{\mathcal{P}}_{\hat{f}}(\mathbf{X}) = \{\hat{\mathcal{P}}(y_0 = 1), \ldots, \hat{\mathcal{P}}(y_n = 1)\}$ for all 700 MC-years. The performance of the entire pipeline is evaluated using two different metrics: predicted target $\hat{\mathbf{y}}$ and predicted probability $\hat{\mathbf{p}}$. These metrics can be compared to determine which is most appropriate for use in a production scenario where only one is part of the metamodel.

Aggregation. The aggregation step serves to mitigate the imprecision of predictions made by the classification model. The LOLE, being an aggregate measure itself, does not require the correct classification of the maximal price indicator (MPI) in every hour, which is a highly challenging task. Instead, the accuracy of the LOLE prediction depends on the precision of the aggregated classification within each MC-year and region.

The aggregation process involves summing the predicted targets $\hat{\mathbf{y}}$ over the MC-year regions:

$$\begin{aligned}\hat{\mathbf{y}}^{\Sigma} &= \begin{bmatrix} \hat{y}^{\Sigma}_{os=1,cy=1982,ma=\mathrm{AT00}} \\ \vdots \\ \hat{y}^{\Sigma}_{os=20,cy=2016,ma=\mathrm{PL00}} \end{bmatrix} \\ &= \begin{bmatrix} \sum_{h\in\mathbb{H}} \hat{y}_{os=1,cy=1982,ma=\mathrm{AT00},h} \\ \vdots \\ \sum_{h\in\mathbb{H}} \hat{y}_{os=20,cy=2016,ma=\mathrm{PL00},h} \end{bmatrix},\end{aligned} \tag{1}$$

with $\hat{\mathbf{y}}^{\Sigma} \in \mathbb{R}^{|MC|\cdot|\mathbb{M}|\times 1}$.

The predicted probabilities are aggregated in the same way, resulting in $\hat{\mathbf{p}}^{\Sigma} \in \mathbb{R}^{|MC|\cdot|\mathbb{M}|\times 1}$. The simulated target is aggregated only over the simulated MC-years MC_{train}, resulting in $\mathbf{y}^{\Sigma}_{\mathrm{train}} \in \mathbb{R}^{|MC_{\mathrm{train}}|\cdot|\mathbb{M}|\times 1}$.

Scaling. A systematic error in the classification model can be corrected by the scaling model. Both systematic underestimation and overestimation can be captured. However, the error must be consistent. This means that the aggregated prediction $\hat{\mathbf{y}}^{\Sigma}$ or probability $\hat{\mathbf{p}}^{\Sigma}$ must exhibit a linear relationship with the aggregated target $\mathbf{y}^{\Sigma}$.

To train the scaling model, the simulated and aggregated targets from the training set $\mathbf{y}^{\Sigma}_{\mathrm{train}}$ are used together with their corresponding predictions $\hat{\mathbf{y}}^{\Sigma}_{\mathrm{train}}$ and $\hat{\mathbf{p}}^{\Sigma}_{\mathrm{train}}$ to fit a linear regression. The regression aims to approximate $\hat{g} \approx g : \hat{\mathbf{y}}^{\Sigma}_{\mathrm{train}} \rightarrow \mathbf{y}^{\Sigma}_{\mathrm{train}}$ and $\hat{h} \approx h : \hat{\mathbf{p}}^{\Sigma}_{\mathrm{train}} \rightarrow \mathbf{y}^{\Sigma}_{\mathrm{train}}$.

Since the linear regression model may produce negative predictions, RELU is used.

Given that LOLE is a non-negative quantity, this function ensures that all predictions are clipped at zero.

The fitted regression models $\hat{g}$ and $\hat{h}$ are used to obtain the scaled predictions:

$$\widehat{\hat{\mathbf{y}}^{\Sigma}} = \hat{g}(\hat{\mathbf{y}}^{\Sigma})^{+} \text{ and } \widehat{\hat{\mathbf{p}}^{\Sigma}} = \hat{h}(\hat{\mathbf{p}}^{\Sigma})^{+}. \tag{2}$$

4 Iterative Guided Convergence for Early Stopping

All MC-years are divided into two sets. Those not yet simulated (i.e., without a known $\mathbf{z}$) are placed in the MC_{unsim} set, while those already simulated with an associated $\mathbf{z}$ are placed in MC_{sim}. Initially, all MC-years are part of MC_{unsim} ($MC_{\text{unsim}} = MC$), while $MC_{\text{sim}} = \emptyset$.

In each iteration, a subset $MC_{\text{batch}} \subset MC_{\text{unsim}}$ is selected and simulated to obtain the corresponding targets $\mathbf{z}_{\text{batch}}$. This batch is appended to the existing simulated set, updating both MC_{sim} and $\mathbf{z}_{\text{sim}}$.

Next, the convergence of $\mathbf{z}_{\text{sim}}$ is evaluated. Convergence is considered achieved when the difference between the simulated and predicted final LOLE remains consistently below a predefined threshold. Once convergence is achieved, the process terminates and the final LOLE is derived from $\mathbf{z}_{\text{sim}}$. If not, the process repeats with the selection of a new batch. This cycle of selection, simulation, and convergence-checking continues.

Figure 2 illustrates the overall process.

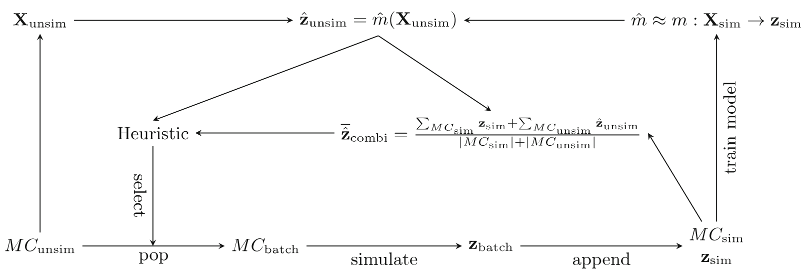

Fig. 2. Proposed iterative guided convergence process for early stopping.

If the iterative process is executed with random selection of MC-years for MC_{batch}, its potential benefits are not fully exploited. To maximize efficiency, priority should be given to simulate the MC-years with the highest number of scarcity occurrences, ensuring that these are processed early in the sequence. A heuristic for selecting MC-years for MC_{batch} in each iteration enables the selection of highly informative samples early on, thus guiding the simulation order and facilitating faster convergence. Fast convergence implies that only a subset of MC-years must be simulated to obtain sufficiently accurate approximations, thereby allowing early termination of the process.

The success of this heuristic depends on access to information about the unsimulated MC-years. This information is derived from a metamodel trained on the already simulated MC-years. The metamodel is retrained in every iteration to improve its predictive performance. In this way, the model can actively select the most informative samples, those that deviate the most or carry the most information, for its own retraining. This configuration, in which the metamodel selects its own training data, represents a form of active learning [27,28].

The metamodel $\hat{m}$, trained on simulated data $\mathbf{X}_{\text{sim}}$ and $\mathbf{z}_{\text{sim}}$, is used to make predictions $\hat{\mathbf{z}}$ for the unsimulated data: $\hat{\mathbf{z}}_{\text{unsim}} = \hat{m}(\mathbf{X}_{\text{unsim}})$.

The input data $\mathbf{X}$ contains a different number of rows than $\hat{\mathbf{z}}$. Specifically, $\mathbf{X}$ consists of $|MC| \cdot |\mathbb{M}| \cdot |\mathbb{H}|$ rows, while $\hat{\mathbf{z}}$ is aggregated over hours, resulting in $|MC| \cdot |\mathbb{M}|$ rows.

To approximate the LOLE per MA, the simulated targets $\mathbf{z}_{\text{sim}}$ and predicted targets $\hat{\mathbf{z}}_{\text{unsim}}$ are used. The average LOLE per MA for the simulated set is:

$$\overline{\mathbf{z}}_{\text{sim}} = \frac{\sum_{MC_{\text{sim}}} \mathbf{z}_{\text{sim}}}{|MC_{\text{sim}}|}, \quad \overline{\mathbf{z}}_{\text{sim}} \in \mathbb{R}^{|\mathbb{M}| \times 1}. \tag{3}$$

The corresponding average for the predicted targets of the unsimulated set can be calculated analog to Eq. (3).

By combining $\overline{\mathbf{z}}_{\text{sim}}$ and $\overline{\hat{\mathbf{z}}}_{\text{unsim}}$, a comprehensive prediction for the final LOLE across all MC-years can be computed:

$$\begin{aligned} \overline{\hat{\mathbf{z}}}_{\text{combi}} &= \overline{\{\mathbf{z}_{\text{sim}}, \hat{\mathbf{z}}_{\text{unsim}}\}} \\ &= \frac{\sum_{MC_{\text{sim}}} \mathbf{z}_{\text{sim}} + \sum_{MC_{\text{unsim}}} \hat{\mathbf{z}}_{\text{unsim}}}{|MC_{\text{sim}}| + |MC_{\text{unsim}}|}, \quad \overline{\hat{\mathbf{z}}}_{\text{combi}} \in \mathbb{R}^{|\mathbb{M}| \times 1}. \end{aligned} \tag{4}$$

The final LOLE approximation $\overline{\hat{\mathbf{z}}}_{\text{combi}}$ can then be compared with the predictions $\hat{\mathbf{z}}$ for the unsimulated MC-year regions. This comparison allows the derivation of a heuristic that selects the next batch of MC-years for simulation in the following iteration.

4.1 Heuristic

The identification of the optimal set of MC-years for simulation in the next iteration involves balancing two objectives:

1. Accelerating the convergence of the LOLE computed from the set of simulated MC-years.
2. Prioritizing the simulation of MC-years with high information content as early as possible.

The initial step determines whether the current LOLE of the simulated MC-years $\overline{\mathbf{z}}_{\text{sim}}$ represents an overestimation or underestimation of the final approximated LOLE $\overline{\hat{\mathbf{z}}}_{\text{combi}}$. Each additional MC-year selected for simulation should ideally reduce the distance between $\overline{\mathbf{z}}_{\text{sim}}$ and $\overline{\hat{\mathbf{z}}}_{\text{combi}}$. This is based on the rationale that $\overline{\hat{\mathbf{z}}}_{\text{combi}}$ serves as an estimate of the final value to which $\overline{\mathbf{z}}_{\text{sim}}$ should converge.

To evaluate the effect of simulating a candidate MC-year mc, the metric $\overline{\hat{\mathbf{z}}}_{MC_{\text{sim}} \cup \{mc\}}$ is introduced. It approximates the new LOLE per MA under the hypothetical assumption that mc is added to the simulation set:

$$\overline{\hat{\mathbf{z}}}_{MC_{\text{sim}} \cup \{mc\}} = \frac{\mathbf{z}_{\text{sim}} + \hat{\mathbf{z}}_{mc}}{|MC_{\text{sim}}| + 1} \tag{5}$$

A binary decision function $d(mc)$ is defined to assess whether simulating mc would move $\overline{\mathbf{z}}_{\text{sim}}$ closer to $\overline{\hat{\mathbf{z}}}_{\text{combi}}$. The function returns 1 if the inclusion of mc reduces the distance between the two values, and 0 otherwise:

$$d(mc) = \begin{cases} 1, & \bar{\mathbf{z}}_{\text{sim}} < \bar{\hat{\mathbf{z}}}_{\text{combi}} \text{ and } \bar{\mathbf{z}}_{\text{sim}} < \bar{\hat{\mathbf{z}}}_{MC_{\text{sim}} \cup \{mc\}} \\ 1, & \bar{\mathbf{z}}_{\text{sim}} > \bar{\hat{\mathbf{z}}}_{\text{combi}} \text{ and } \bar{\mathbf{z}}_{\text{sim}} > \bar{\hat{\mathbf{z}}}_{MC_{\text{sim}} \cup \{mc\}} \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

This formulation ensures that the heuristic promotes both convergence of $\bar{\mathbf{z}}_{\text{sim}}$ toward $\bar{\hat{\mathbf{z}}}_{\text{combi}}$ and the early simulation of outlier MC-years.

To quantify the relative influence of simulating a specific mc on the final LOLE, the following impact metric is defined:

$$\text{impact}(mc) = \left| \frac{\bar{\hat{\mathbf{z}}}_{MC_{\text{sim}} \cup \{mc\}} - \bar{\hat{\mathbf{z}}}_{\text{combi}}}{\bar{\hat{\mathbf{z}}}_{\text{combi}}} \right| \tag{7}$$

Combining the decision function and impact metric, the next MC-year to simulate is chosen as the one that maximizes the product of both:

$$mc_{\text{next}} = \text{argmax}_{mc \in MC_{\text{unsim}}} \left(d(mc) \cdot \text{impact}(mc) \right) \tag{8}$$

This heuristic ensures that extreme and informative MC-years are simulated promptly, thereby accelerating convergence toward the final LOLE estimate.

4.2 Selecting the Size and Content of the First Batch

The size of the initial batch can be chosen relatively small if the metamodel supports refitting-i.e., if it is sufficient to update the model parameters using new data, rather than retraining the model from scratch.

The minimum batch size should correspond to the number of simulations that can be executed in parallel. If the computational overhead of launching new simulations with a specified number of MC-years is negligible, the batch size should ideally match the number of available parallel simulation slots.

While the metamodel is capable of leveraging the iterative process to select its own training MC-years, the initial training set must be selected manually or via random sampling. In this context, domain knowledge can be employed to guide the selection of a representative and informative range of MC-years for initial model training.

Importantly, this selection must be made prior to observing any simulation results and should rely solely on features derived from the input data. One possible strategy for selecting the initial batch involves evaluating remaining capacity and applying techniques from the field of DOE.

5 Experimental Setup

To ensure a robust evaluation of model performance, a k-fold cross-validation(CV) procedure is employed. The data is partitioned into $k = 5$ folds, each consisting of 140 MC-years. Four of the five folds are used for training and validation, while the fifth functions as a holdout set. This holdout set is not involved in any training or validation but is reserved solely for assessing and comparing model performance on a consistent dataset.

This setup allows for an investigation into how the selection of training MC-years influences model performance. If significant variability is observed across the models trained on different folds, it indicates that the choice of training data is crucial for achieving optimal predictive accuracy.

Two versions of the model are implemented:

- A *universal model*, trained on the complete dataset across all MAs. This version applies undersampling, preserving all MPE cases and only 4% of non-MPEs.
- An *ensemble of region-specific models*, trained individually for each MA. These models are trained sequentially because the data are partitioned by region, which eliminates the need for undersampling due to data imbalance.

6 Metamodel Topology Development

In the preliminary phase of evaluating the classification model selection and hyperparameter optimization process, the models that emerged from a comprehensive exploration are presented. These models are subsequently thoroughly assessed for their performance as an ensemble model. The model that demonstrates the best performance is then evaluated embedded in the metamodel pipeline in combination with a scaling model.

6.1 Grid Search on Different Models

The most promising model hyperparameter combinations after training and evaluating over 500 models from four model types are

- two support vector classifiers with polynomial kernels of degree two and seven,
- an multi layer perceptron (MLP) with one hidden layer having 20 neurons, and
- a logistic regression classifier.

6.2 Comparing Model Types as Regional Ensemble Learners

The models that emerged from the grid search are assembled as an ensemble learner, with one specialized model for each MA. The four models are evaluated on the full dataset. The results of the ensemble models for the datasets `t+1` can be seen in Fig. 3. The color and size represent the F_1-score of each regional model. The scores labeled AGG on the x-axis are the scores aggregated over all regions. It is important to note that the aggregated F_1-score is not simply the mean of the regional F_1-scores. Rather, it is calculated using the sum of the true positives, true negatives, false positives, and false negatives over all MAs.

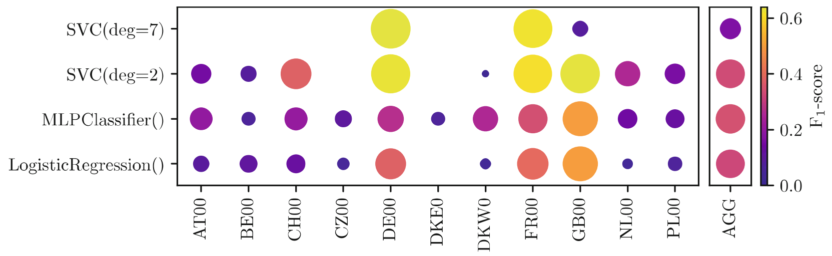

Fig. 3. Holdout F_1-score for regional ensemble learner (`t+1`).

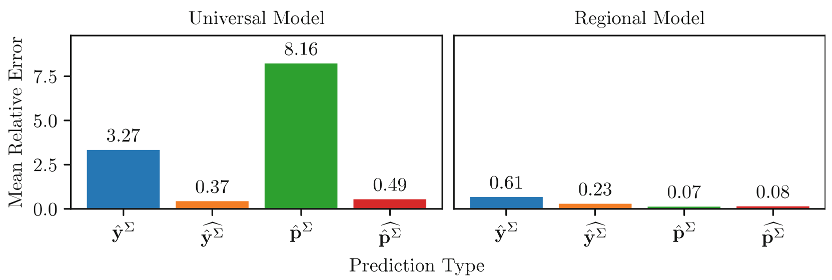

Fig. 4. Relative error using the full metamodels trained on 20% of the data. The final prediction combines the target of the training set and the predicted target or target probability of the validation and holdout set on `t+1`.

The final classification model is a support vector classifier with a polynomial kernel of degree two. Both model variants are evaluated using the k-fold CV approach, resulting in four cross-validation models (CVM 14).

A comparison of the relative error regarding the CVMs in predicting the final LOLE for different configurations of the metamodel pipeline in Fig. 4 shows that scaling effects are more pronounced in the less effective classification model (the universal model). The optimal configuration involves utilizing the regional model and its unscaled probability prediction, $\hat{\mathbf{p}}^{\Sigma}$. However, the differences between the unscaled and scaled probability predictions for the regional model are minimal. The universal model without scaling demonstrates a relative error of more than 300% or 800%, respectively. Such inaccuracy renders the universal classification model without scaling unsuitable for practical application.

Figure 5 and 6 present the relative error of the final LOLE predictions per MA.

Figure 5 illustrates the performance of the regional ensemble classification model trained via grid search and evaluated using unscaled predicted targets ($\hat{\mathbf{y}}^{\Sigma}$).

In contrast, Fig. 6 shows the results of the region-specific ensemble learner that predicts the probability of an MPE, followed by a scaling model to estimate the final LOLE ($\widehat{\hat{\mathbf{p}}^{\Sigma}}$). This approach achieves an average relative error of less than 10% across the four CVMs and regions, while using only 20% of the data.

This result demonstrates that even with limited simulated data, the metamodel can deliver highly accurate LOLE estimates when appropriate modeling strategies are used, such as regional specialization and scaling.

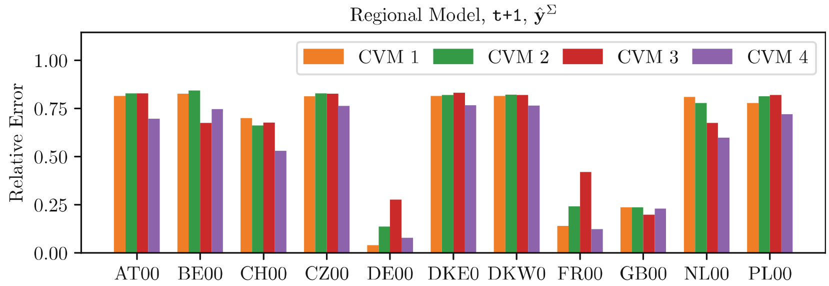

Fig. 5. Relative error of the final prediction using $\hat{\mathbf{y}}^{\Sigma}$ (combining the target of the training set $\mathbf{y}^{\Sigma}$ with the unscaled predicted target $\hat{\mathbf{y}}^{\Sigma}$ of the validation and holdout sets) for the CVMs on `t+1`.

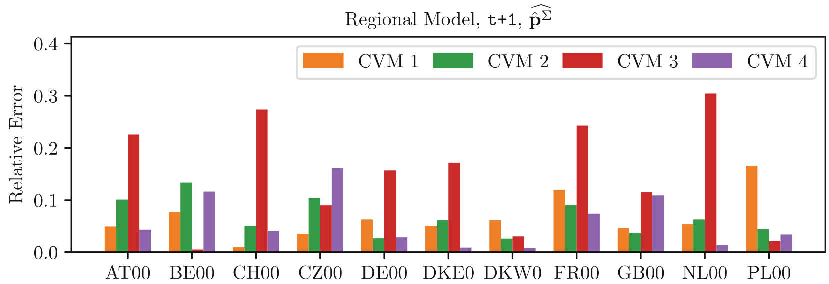

Fig. 6. Relative error of the final prediction using $\widehat{\mathbf{p}}^{\Sigma}$ (combining the target of the training set $\mathbf{y}^{\Sigma}$ with the scaled predicted probability $\widehat{\mathbf{p}}^{\Sigma}$ of the validation and holdout sets) of the CVMs on `t+1`.

6.3 Evaluation of Iterative Guided Convergence for Early Stopping

To evaluate the effectiveness of the algorithm outlined in Sect. 4, the convergence of the LOLE is analyzed by sorting the MC-years chronologically according to CY and OS.

Figures 7 and 8 compare the guided convergence performance of different prediction metrics for universal and regional models. For the universal model, it is evident that scaling substantially improves prediction accuracy. In contrast, all metrics in the regional model perform adequately, except for $\widehat{\mathbf{y}}^{\Sigma}$, which shows notable deviation.

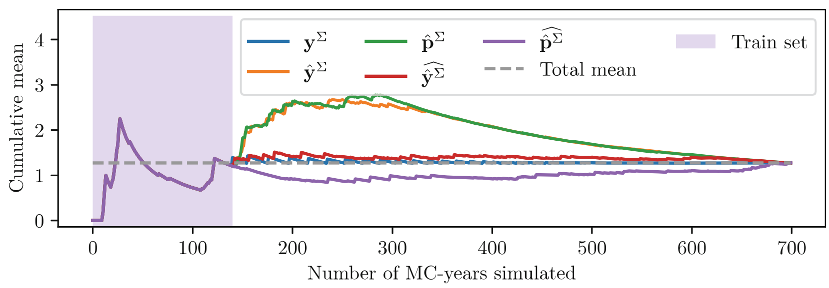

Fig. 7. Cumulative and total mean; MC-years sorted by heuristic for `DE00`; comparison of different prediction metrics for the universal model (CVM 1) on `t+1`.

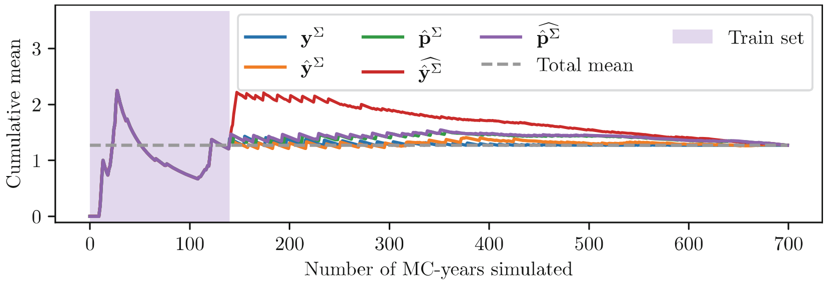

Fig. 8. Cumulative and total mean; MC-years sorted by heuristic for `DE00`; comparison of different prediction metrics for the regional model (CVM 1) on `t+1`.

These results confirm that the convergence guidance algorithm performs reliably in conjunction with real metamodels. Nonetheless, further exploration is needed to determine whether iterative guided convergence can reduce the required training set to less than 20% by allowing the metamodel to select its training samples dynamically.

6.4 Computational Complexity

The reduction in computational effort is directly related to the number of MC-years that must be simulated and the training time required for the metamodels.

Using 140 MC-years as an upper bound for training, the minimum achievable computational savings can be estimated. For model training, the universal and regional metamodels must be considered separately. The universal model requires only one training procedure, while the regional model necessitates the sequential training of 11 models (one per MA).). Training time for the regional model increases proportionally unless executed in parallel.

Simulating all 700 MC-years requires roughly 144 h (six days). By reducing the simulation scope to 140 MC-years (20%), the total simulation time decreases to about 28.8 h (1.2 days). Model training on this reduced dataset takes approximately 20 min for the universal model and 3.3 h for the regional models. Inference on the remaining 560 MC-years requires an additional 1 h for either model type.

The computational time saved is then:

$$\frac{144h - (28.8h + 0.33h + 1h)}{144h} \approx 79\% \tag{9}$$

for the universal model, and

$$\frac{144h - (28.8h + 3.3h + 1h)}{144h} \approx 77\% \tag{10}$$

for the regional model.

These savings highlight the substantial efficiency gains achieved through guided convergence and metamodeling.

7 Conclusion and Discussion

The most effective and robust pipeline configuration consists of three components: a regional ensemble classification model predicting the probabilities of MPEs as the output and an aggregation step. We could also demonstrate that incorporating our linear regression as an additional step in the metamodelling pipeline can significantly improve predictions, especially for the universal model. While using squared loss in the scaling step may lead to a higher MAE, it is preferred because it better penalizes large errors, ensuring that extreme and critical scenarios are not overlooked.

The iterative guided convergence method introduced in this work demonstrates strong potential, particularly as a framework for active learning. By prioritizing the simulation of MC-years that diverge most from already simulated ones, the method effectively identifies data points with high information content. This makes the approach ideal for generating meaningful training data in an efficient and targeted manner. It is important to emphasize that metamodeling does not replace traditional MC simulations. Instead, it serves as a complementary tool that can reduce the number of necessary simulations. Since the metamodel must be trained on already simulated data, its effectiveness is contingent on the availability and quality of those initial simulations. Future studies could improve data preprocessing by incorporating techniques such as principal component analysis (PCA). PCA has the potential to reduce data dimensionality while retaining most of the variance, thereby speeding up model training and decreasing data storage requirements with minimal loss in accuracy.

Furthermore, temporal features, including rolling window aggregations, have the potential to be further developed in order to quantify the impact of renewable energy droughts on system adequacy. It can be proposed that the optimisation of the window length may yield further enhancements, with the objective of maximising correlation with system stress events.

The classification component of the metamodel could be enhanced by exploring advanced ML models, including deep neural networks. The scaling model also offers room for improvement. Although a linear model was used in this work, more complex models such as polynomial regression could better capture nonlinear relationships. Nevertheless, the risk of overfitting must be given due consideration. Alternatively, the adoption of a regional ensemble approach for scaling, similar to the classification stage, might yield performance gains without sacrificing generalizability.

References

1. Fernandez, R.M.: Conflicting energy policy priorities in EU energy governance. J. Environ. Stud. Sci. **8**(3), 239–248 (2018). ISSN 2190-6483, 2190-6491. https://doi.org/10.1007/s13412-018-0499-0. Accessed 18 Mar 2025
2. Marhold, A.-A.: Towards a 'security-centred' energy transition: balancing the European Union's ambitions and geopolitical realities. J. Int. Econ. Law **26**(4), 756–769 (2024). ISSN 1369-3034, 1464-3758. https://doi.org/10.1093/jiel/jgad043. https://academic.oup.com/jiel/article/26/4/756/7479918. Accessed 18 Feb 2025
3. Methodology for the European resource adequacy assessment. ACER, 2 October 2020. https://www.acer.europa.eu/sites/default/files/documents/Individual%20Decisions_annex/ACER%20Decision%2024-2020%20on%20ERAA%20-%20Annex%20I_1.pdf. Accessed 23 Oct 2024
4. Köhnen, C.S., et al.: The potential of deep learning to reduce complexity in energy system modeling. Int. J. Energy Res. **46**(4) (2022), 4550–4571. ISSN 0363-907X, 1099-114X. https://doi.org/10.1002/er.7448. https://onlinelibrary.wiley.com/doi/10.1002/er.7448. Accessed 16 Oct 2024
5. Starke, A.R., Da Silva, A.K.: A review on the applicability of machine learning techniques to the metamodeling of energy systems. Numer. Heat Transf. Part B Fundam., 1–30 (2023). ISSN 1040-7790, 1521-0626. https://doi.org/10.1080/10407790.2023.2280208. https://www.tandfonline.com/doi/full/10.1080/10407790.2023.2280208. Accessed 16 Oct 2024
6. Rahman, Md.M., et al.: Prospective methodologies in hybrid renewable energy systems for energy prediction using artificial neural networks. Sustainability **13**(4), 2393 (2021). ISSN 2071-1050. https://doi.org/10.3390/su13042393. https://www.mdpi.com/2071-1050/13/4/2393 (visited on 03/12/2025)
7. Donti, P.L., Zico Kolter, J.: Machine learning for sustainable energy systems. Ann. Rev. Environ. Res. **46**(1), 719–747 (2021). ISSN 1543-5938, 1545-2050. https://doi.org/10.1146/annurev-environ-020220-061831. https://www.annualreviews.org/doi/10.1146/annurev-environ-020220-061831. Accessed 16 Oct 2024
8. Duchesne, L., Karangelos, E., Wehenkel, L.: Recent developments in machine learning for energy systems reliability management. Proc. IEEE **108**(9), 1656–1676 (2020). ISSN 0018-9219, 1558-2256. https://doi.org/10.1109/JPROC.2020.2988715. https://ieeexplore.ieee.org/document/9091534/. Accessed 12 Mar 2025
9. Sun, Y., et al.: Local feature sufficiency exploration for predicting security-constrained generation dispatch in multi-area power systems. In: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, December 2018, pp. 1283–1289. IEEE (2018). ISSN 9781538668054. https://doi.org/10.1109/ICMLA.2018.00208. https://ieeexplore.ieee.org/document/8614233/. Accessed 12 Mar 2025
10. Su, H., et al.: An integrated, systematic data-driven supply-demand side management method for smart integrated energy systems. Energy **235**, 121416 (2021). ISSN 03605442. https://doi.org/10.1016/j.energy.2021.121416. https://linkinghub.elsevier.com/retrieve/pii/S0360544221016649. Accessed 12 Mar 2025
11. Leonori, S., et al.: Microgrid energy management systems design by computational intelligence techniques. Appl. Energy **277**, 115524 (2020). ISSN 03062619. https://doi.org/10.1016/j.apenergy.2020.115524. http://linkinghub.elsevier.com/retrieve/pii/S0306261920310369. Accessed 12 Mar 2025
12. Huang, Y., Xu, Q., Lin, G.: Congestion risk-averse stochastic unit commitment with transmission reserves in wind-thermal power systems. Appl. Sci. **8**(10), 1726

(2018). ISSN 2076-3417. https://doi.org/10.3390/app8101726. http://www.mdpi.com/2076-3417/8/10/1726. Accessed 12 Mar 2025
13. Darville, J., Curia, J., Celik, N.: Microgrid operational planning using a hybrid neural network with resource-aware scenario selection. Simul. Model. Pract. Theor. **119**, 102583 (2022). ISSN 1569190X. https://doi.org/10.1016/j.simpat.2022.102583. http://linkinghub.elsevier.com/retrieve/pii/S1569190X22000715. Accessed 23 Oct 2024
14. Bogensperger, A.J., Fabel, Y., Ferstl, J.: Accelerating energy-economic simulation models via machine learning-based emulation and time series aggregation. Energies **15**(3), 1239 (2022). ISSN 1996-1073. https://doi.org/10.3390/en15031239. http://www.mdpi.com/1996-1073/15/3/1239. Accessed 23 Oct 2024
15. Hoffmann, M., et al.: A review on time series aggregation methods for energy system models. Energies **13**(3), 641 (2020). ISSN 1996-1073. https://doi.org/10.3390/en13030641. http://www.mdpi.com/1996-1073/13/3/641. Accessed 24 Oct 2024
16. Perera, A.T.D., et al.: Machine learning methods to assist energy system optimization. Appl. Energy **243**, 191–205 (2019). ISSN 03062619. https://doi.org/10.1016/j.apenergy.2019.03.202. http://linkinghub.elsevier.com/retrieve/pii/S030626191930618X. Accessed 12 Mar 2025
17. Prina, M.G., et al.: Machine learning as a surrogate model for EnergyPLAN: speeding up energy system optimization at the country level. Energy **307**, 132735 (2024). ISSN 03605442. https://doi.org/10.1016/j.energy.2024.132735. http://linkinghub.elsevier.com/retrieve/pii/S036054422402509X. Accessed 23 Oct 2024
18. Danish, M.S.S.: A framework for modeling and optimization of data-driven energy systems using machine learning. IEEE Trans. Artif. Intell. **5**(5), 2434–2443 (2024). ISSN 2691-4581. https://doi.org/10.1109/TAI.2023.3322395. http://ieeexplore.ieee.org/document/10273612/. Accessed 23 Oct 2024
19. Kim, M.J., et al.: Neural-network-based optimization for economic dispatch of combined heat and power systems. Appl. Energy **265**, 114785 (2020). ISSN 03062619. https://doi.org/10.1016/j.apenergy.2020.114785. http://linkinghub.elsevier.com/retrieve/pii/S030626192030297X. Accessed 23 Oct 2024
20. Nolting, L., et al.: Can energy system modeling benefit from artificial neural networks? Application of two-stage metamodels to reduce computation of security of supply assessments. Comput. Ind. Eng. **142**, 106334 (2020). ISSN 03608352. https://doi.org/10.1016/j.cie.2020.106334. http://linkinghub.elsevier.com/retrieve/pii/S0360835220300681. Accessed 21 Oct 2024
21. Priesmann, J., Nolting, L., Praktiknjo, A.: Are complex energy system models more accurate? An intra-model comparison of power system optimization models. Appl. Energy **255**, 113783 (2019). ISSN 03062619. https://doi.org/10.1016/j.apenergy.2019.113783. http://linkinghub.elsevier.com/retrieve/pii/S0306261919314709. Accessed 24 Oct 2024
22. Nolting, L., Praktiknjo, A.: The complexity dilemma – insights from security of electricity supply assessments. Energy **241**, 122522 (2022). ISSN 03605442. https://doi.org/10.1016/j.energy.2021.122522. http://linkinghub.elsevier.com/retrieve/pii/S0360544221027717. Accessed 23 Oct 2024
23. Billinton, R., Allan, R.N.: Reliability Evaluation of Power Systems. Springer, Boston (1996). ISBN 9781489918628 9781489918604. https://doi.org/10.1007/978-1-4899-1860-4. http://link.springer.com/10.1007/978-1-4899-1860-4. Accessed 31 Oct 2024

24. Münch, J., et al.: Uplifting the complexity of analysis for probabilistic security of electricity supply assessments using artificial neural networks. Energy AI **17**, 100401 (2024). ISSN 26665468. https://doi.org/10.1016/j.egyai.2024.100401. http://linkinghub.elsevier.com/retrieve/pii/S2666546824000673. Accessed 23 Oct 2024
25. Gonzato, S., Bruninx, K., Delarue, E.: The effect of short term storage operation on resource adequacy. Sustain. Energy Grids Netw. **34**, 101005 (2023)
26. Fraunholz, C., et al.: Demand curtailment allocation in interconnected electricity markets. Appl. Energy **377**, 124679 (2025)
27. Settles, B.: Active Learning. Synthesis Lectures on Artificial Intelligence and Machine Learning. Springer, Cham (2012). 9783031004322. https://doi.org/10.1007/978-3-031-01560-1. http://link.springer.com/10.1007/978-3-031-01560-1. Accessed 25 Oct 2024
28. Settles, B.: Active learning literature survey (2009). http://www.semanticscholar.org/paper/Active-Learning-Literature-Survey-Settles/818826f356444f3daa3447755bf63f171f39ec47. Accessed 25 Oct 2024