Blending Surrogates for Energy System Modeling

Hussein Genena, Anurag Mohapatra, Beneharo Reveron Baecker, Thomas Hamacher
Technical University of Munich, Germany
{*hussein.genena, anurag.mohapatra, beneharo.reveron-baecker, thomas.hamacher*}*@tum.de*

***Abstract*—Surrogate modeling offers a computationally efficient alternative to energy system modeling (ESM) frameworks for grid planning. However, constructing surrogate models for large distribution grids (DG) comprising multiple sub-grids remains challenging due to the limited availability of ground-truth data. This paper investigates methods for developing surrogates for large DG that predict the net hourly electricity demand, assuming that surrogate models of the individual, smaller sub-grids are already available and limited ground-truth data availability. The explored methods rely on weight averaging, hourly weighted blending using linear regression, neural network-based aggregation, and synthetic data creation, respectively. Among these, the synthetic data approach demonstrates the highest accuracy, effectively preserving the temporal shape of the net demand while addressing the data scarcity challenge. Further evaluation under different data conditions confirms the consistency of its performance, while an additional experiment with a larger grid ensemble validates the scalability of the proposed approach. The work proves that surrogates of sub-grids can be reliably blended with little to no ground truth originating from the larger composite grid.**

***Index Terms*—Machine Learning, Energy System Modeling, Active Distribution Grids, LSTM, Transformers**

I. Introduction

Germany's goal of achieving net greenhouse gas neutrality by 2045 [1] demands massive infrastructure changes to the power grids. Specifically, this transition will heavily impact DG, as most renewable energy sources are decentralized and connected at the distribution level. In fact, the cumulative distribution network investments between 2024 and 2045 are expected to exceed 200 billion euros [2], with corresponding upgrades required in upstream and transmission grids to handle increased power fluctuations.

Detailed expansion planning is necessary to avoid inflated costs for distribution and transmission system operators (DSO and TSO) while maintaining a reliable electricity supply for consumers, under the ambitious decarbonization objectives. Traditional energy system planning and transmission grid studies have often substituted DG as a single net demand time series [3]. However, the operation of sector-coupled DG is significantly impacted by the increasing penetration of renewable energy resources, the wider deployment of flexibility measures, and evolving energy market dynamics. Consequently, the net demand at the DG's top node can no longer be accurately estimated by merely combining inelastic demand and generation time series. Instead, detailed DG-level models that explicitly consider flexibility options are required to properly assess their impact on upstream grids and enable their integration into system-level planning.

With the rapid evolution of the power system landscape driven by the electrification of heating and mobility, including their associated flexibility potentials, the modeling of sector-coupled DG assets has become substantially more complex. Simultaneously, rising demand and stricter technical constraints require more detailed grid representations. This higher level of model detail significantly increases the time and resources required for data acquisition, a process further hindered by data privacy and security constraints that restrict access to high-resolution datasets. Moreover, the resulting growth in model complexity amplifies computational demands and can render large-scale problems computationally intractable.

To circumvent these problems, machine learning (ML) surrogate models have emerged as a promising approach. A key advantage of surrogate models for grid planners lies in their resource-efficient design, as they are developed to generate a specific set of outputs from a predefined set of input variables without performing full system optimization.

A. Research Gap

In [4], we developed long short-term memory- (LSTM) / transformer-based surrogate models capable of predicting the hourly net demand of a DG, with significant computational savings compared to ESMs. These surrogates were trained using specific features derived from 27 ESM optimization runs, each representing a distinct distributed energy resources' (DER) penetration scenario, and enhanced through data augmentation techniques to better capture temporal variability. This work was later extended in [5] to additionally predict the bus voltages within a single DG using graph neural networks.

However, these studies revealed a key limitation that hinders the scalability of these surrogate models. To ensure accurate and reliable predictions across diverse operating conditions, surrogate models must be trained on a large number of datasets, typically generated using ESMs, that cover a wide operational range. For large distribution grids containing multiple sub-grids, this training data acquisition stage becomes a significant bottleneck. As the grid size increases, the availability of ground-truth datasets is expected to diminish, since coordinating data across interconnected energy networks and collecting high-resolution data from numerous assets becomes increasingly difficult.

A potential solution to this problem could be to aggregate the predictions of the surrogate models of the individual sub-grids, as illustrated on the left-hand side of Fig. 1. As supported by CERRE's assessment of DSOs' limited interaction with other energy networks [6], typically little to no energy

979-8-3315-8138-1

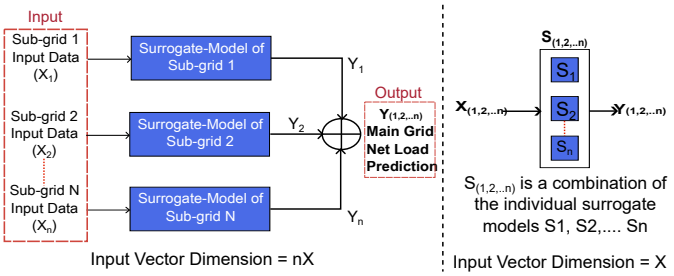

Fig. 1. Reduction of input data dimensionality with a combined model

exchange occurs between sub-grids at the medium- and low-voltage DG levels. Therefore, the sub-grids' net load behaviors can be modeled independently and later combined at a higher level of aggregation. However, this approach is cumbersome and impractical due to the high dimensionality of the input data (nX datasets) and the constant dependence on all sub-grid surrogates for each prediction. Moreover, it is unrealistic to assume that input data on the individual sub-grid level would be available whenever a prediction for the main grid is required. Realistically, input data solely for the main grid would be accessible. This aggregated data should be easier to obtain (e.g., using statistical sources or analytical tools) compared with collecting coordinated, detailed data for each individual sub-grid. Hence, this study focuses on tackling the following research gaps:

- How to build scalable surrogate models for large DGs, when only aggregated main-grid-level input data is available?
- How can such models be built effectively despite the limited availability of ground-truth output data for the large grid?

B. Contribution

This paper explores different methods for merging surrogate models of sub-grids to construct a surrogate for large DGs. Each method aims to predict the net demand at the main-grid level, based on input data at this level, but without access to any sub-grid data.

One of the tested methods, using synthetic datasets generated from the sub-grid surrogates, proves most effective in predicting the net demand, achieving high accuracy despite limited ground-truth data availability for the main grid.

II. Methodology

This section describes the different surrogate model merging methods investigated in this study. Before presenting these methods, the key assumptions and system configurations used in this study are outlined.

A. Assumptions

We use three grids to demonstrate and evaluate the proposed model merging techniques. Two of them, grid A and grid B, are real DGs from the Forchheim region in Germany, representing a rural and a village grid, respectively. These serve as the sub-grids, whose surrogate models are assumed to be readily available and built using the framework in [4].

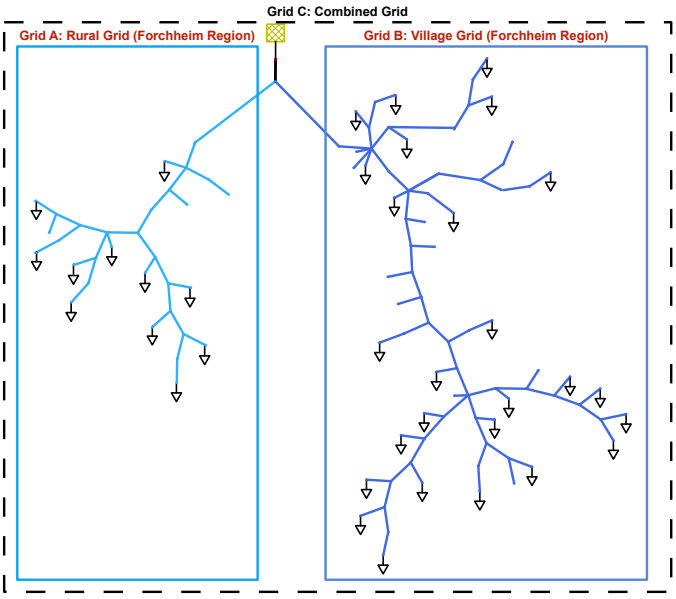

Fig. 2. Structure of Grid C, composed of sub-grids A and B

We create a third grid, grid C, which is an imaginary grid that serves as the main DG in this study. Grid C is composed of grids A and B, as illustrated in Fig. 2, and represents the system for which a new surrogate model is to be developed.

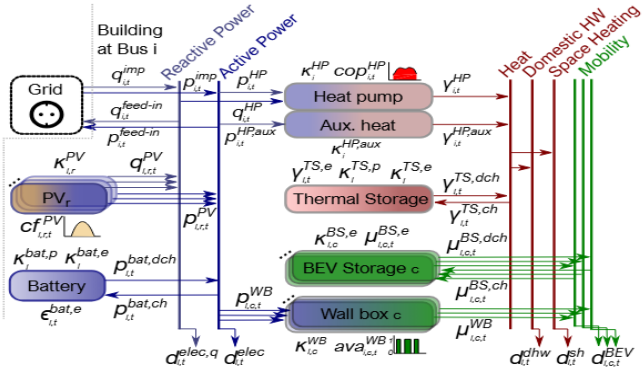

Fig. 3. Urbs building reference energy systems

The ground-truth datasets for Grid C are obtained by optimizing it using the ESM *urbs* [7], a tool used for unit commitment and capacity expansion planning. The corresponding reference energy system is shown in Fig. 3. Grid C is optimized under the $Coor\_Flex+$ scenario from [8], which closely mirrors the current German low-voltage digitalization landscape, characterized by home energy management systems that maximize PV self-consumption with minimal grid-side coordination. This setup also reflects realistic behavior, where households install DER capacities to optimize individual benefits rather than considering the holistic operation of the grid. To reduce computational effort, the time-series aggregation module *tsam* [9] is used to compress the annual grid data into six representative weeks. These same six weeks are consistently used across all experiments to ensure fair comparison and evaluation. However, even with the time savings provided by *tsam*, optimization may still be infeasible for very large DGs. Therefore, in this paper, we also analyze techniques where no

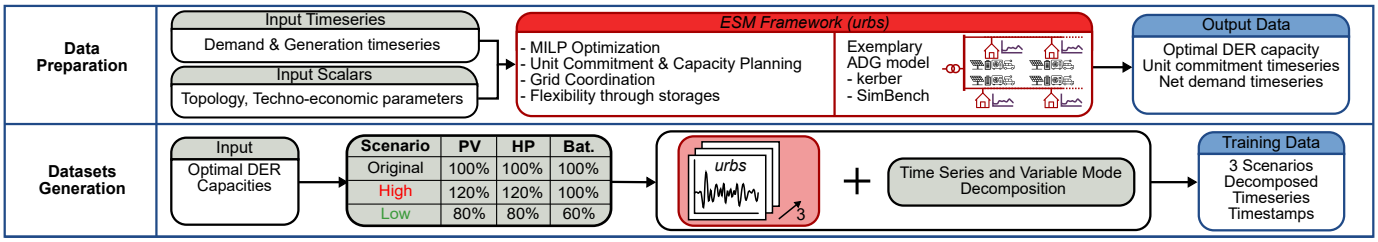

Fig. 4. Schematic overview of the datasets generation process

ground-truth datasets are available during the training phase.

Three datasets are assumed available for Grid C: one real training dataset representing the original 100% DER scenario, and two testing datasets, *high* and *low*. As shown in Fig. 4, Grid C is first optimized to generate the 100% DER dataset. The testing datasets are then produced by re-optimizing the grid after varying the DER ratios. These datasets are also augmented according to the *TS-VMD5* mode used in [4].

B. Model Blending Techniques

This study explored several surrogate model blending techniques, inspired from the literature. The method that successfully addressed both research gaps **relies on hybrid and synthetic datasets**, while the others, summarized in Table I,

TABLE I
DESCRIPTION OF MODEL MERGING METHODOLOGIES

Methodology	Initial Data Requirements
Adaptive Weight Averaging	• Inspired by model fusion in [10] • Linearly combines weights of models A and B to construct grid C's surrogate model • Uses weighting factors based on the sub-grids' net annual demands, instead of simple averaging
Regression Driven Surrogate Model Aggregation	• Inspired by ensemble learning in [11] • Passes grid C data to surrogates A and B • Outputs are multiplied by periodic aggregation factors • Sum of these results is the output of grid C
Hierarchical Merging with Linear Combination and LSTM Fusion	• Passes grid C data to surrogates A and B • Outputs passed to a weighted aggregation layer • This output is fed into an LSTM network

only partially resolved the first gap. Under limited ground-truth availability, these alternative methods fail to produce accurate results due to factors such as normalization mismatch, model inflexibility, and discrepancies in learned weights. Hence, the remainder of this study focuses on the method based on synthetic datasets.

Model Building Using Hybrid and Synthetic Datasets: Under the assumption of uncoordinated DG operation and negligible sub-grid energy exchange, the net demand at Grid C's top node at hour i can be expressed as:

$$d_i^C = d_i^A + d_i^B \tag{1}$$

where d_i^C, d_i^A, and d_i^B are the net demands of grids C, A, and B, respectively, at hour i. This linear relationship motivates the generation of synthetic datasets by combining the outputs of the surrogate models of grids A and B.

This method comprises three distinct steps, as described in Fig. 5.

- **Synchronous datasets** refer to original sub-grids' datasets collected under identical conditions (e.g., same year, DER scenario, and using the same ESOM), allowing corresponding scenario files from grids A and B to be combined directly to create a synthetic scenario for Grid C. However, such conditions are rarely guaranteed in practice, imposing stringent data requirements.
- **Asynchronous datasets** refer to datasets that could originate from different sources, time periods, or under varying scenario conditions. Consequently, when generating synthetic datasets for Grid C, the resulting scenarios may combine sub-grid data from differing conditions.

Fig. 6 shows how synchronous and asynchronous datasets have been implemented for grid C in this study. Accordingly, multiple implementations were designed to handle data synchronicity and were extended to address limited ground-truth availability, as detailed below:

1) **Hybrid** (**Synchronous** Synthetic + Real) Datasets
2) **Hybrid** (**Asynchronous** Synthetic + Real) Datasets
3) **Synchronous Synthetic** Datasets
4) **Asynchronous Synthetic** Datasets

The specific requirements and corresponding testing datasets for each implementation are listed in Table II. In implementations 1 and 2, a custom loss function is applied, where a weighting parameter λ, set to 2, penalizes errors on real data more heavily. Additionally, to evaluate the effect of dataset size on these two hybrid approaches, two models were built using 27 and 140 synthetic datasets, respectively.

III. RESULTS

A. Evaluation of Hybrid and Synthetic Datasets Approach

The first implementation, which required the most data, achieves strong results in both testing scenarios. As shown in Table III, peak demand and feed-in errors remain below 5%, losses under 0.07, and R^2 scores above 0.95, demonstrating the model's capability of preserving the temporal shape of grid C's net demand. Increasing the number of synthetic datasets slightly reduces losses but has a minimal effect on peak errors. The high-scenario feed-in mismatch error observed in the second model arises primarily from time-series aggregation effects rather than limitations of the model-building technique. Fig. 7 reveals that the actual net feed-in at hours 470 and 639

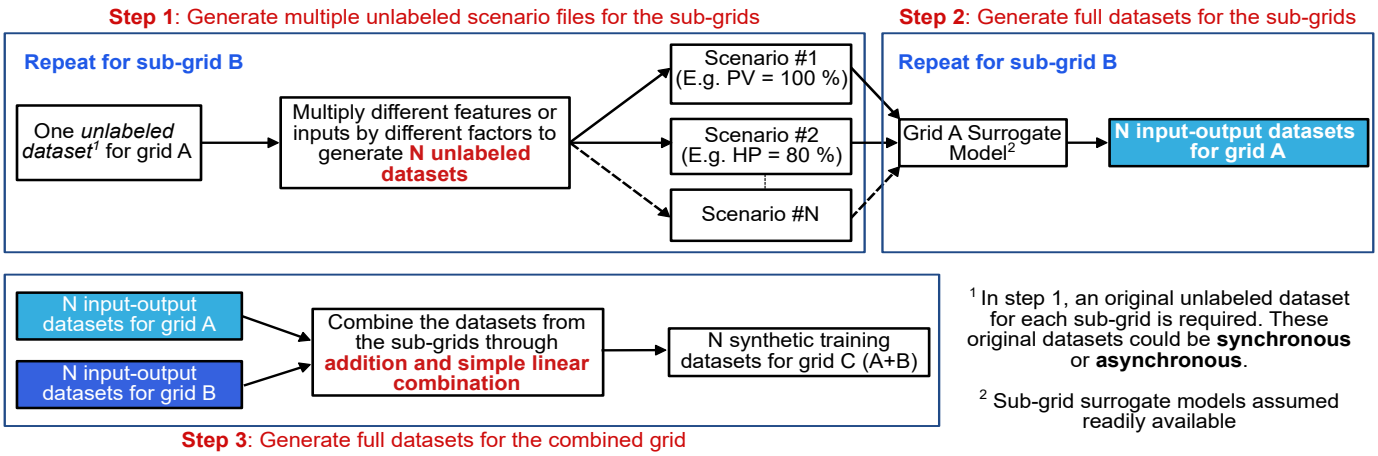

Fig. 5. Sequence for creating synthetic datasets for grid C

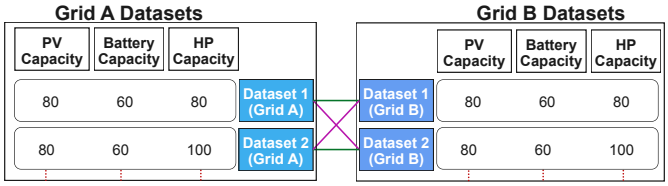

Fig. 6. Asynchronous ■ and synchronous ■ datasets creation

TABLE II
DATA REQUIREMENTS AND TESTING DATASETS FOR THE DIFFERENT IMPLEMENTATIONS

Implementation	Initial Data Requirements
Implementation 1 Hybrid Datasets (Synchronous Synthetic + Real)	Surrogate model for grid A
	Surrogate model for grid B
	Original training dataset for grid C
	Unlabeled dataset for grid A
	Unlabeled dataset for grid B under similar conditions and for the same time period
Implementation 2 Hybrid Datasets (Asynchronous Synthetic + Real)	Surrogate model for grid A
	Surrogate model for grid B
	Original training dataset for grid C
	Random unlabeled dataset for grid A
	Random unlabeled dataset for grid B
Implementation 3 Synthetic Datasets (Synchronous)	Surrogate model for grid A
	Surrogate model for grid B
	Unlabeled dataset for grid A
	Unlabeled dataset for grid B under similar conditions and for the same time period
Implementation 4 Synthetic Datasets (Asynchronous)	Surrogate model for grid A
	Surrogate model for grid B
	Random unlabeled dataset for grid A
	Random unlabeled dataset for grid B
Implementation	**Testing Datasets**
All Implementations	High (PV: 120% , Bat.: 100%, HP: 120%)
	Low (PV: 80% , Bat.: 60%, HP: 80%)

differs by only 2.2%. The model accurately predicts high feed-in values at both times but slightly over-predicts the magnitude at hour 639, leading to the observed peak mismatch.

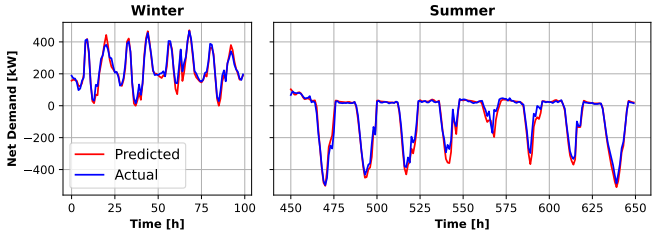

Fig. 7. Actual vs. predicted net demand for the high scenario on representative winter and summer days, using 141 datasets (140 synchronous + 1 real)

In implementation 2, performance strongly depends on the number of synthetic datasets. Using 28 datasets leads to weaker accuracy, especially in the low scenario, whereas using 141 datasets results in substantial improvements. This weaker performance with 28 datasets can be explained by the reduced variability in the training data. In the current implementation, a smaller number of datasets increases the likelihood of the final model being heavily overfit to a narrow range of input features. Unlike the first approach, this method relaxes the need for synchronous datasets, easing data requirements provided sufficient asynchronous datasets are available.

Implementations 3 and 4, which require only synthetic datasets, also perform well. Both accurately track the net demand, and although implementation 4 shows a slightly higher feed-in error (6.36%) in the low scenario, the minimal data requirements could make this trade-off acceptable. Notably, these implementations eliminate the need for ground-truth data or computationally expensive ESM optimization for large grids, representing a crucial advantage that makes the practical deployment of surrogate models for larger grids more feasible.

B. Scalability Assessment of the Hybrid and Synthetic Dataset Approach

To evaluate the scalability of the hybrid and synthetic dataset approach, a larger, imaginary grid, **grid D**, was created, comprising two instances of grid A and three instances of grid B as sub-grids. As shown in Table III, the model maintains

TABLE III
EVALUATION OF MODEL BUILDING FOR GRID C AND D USING HYBRID AND SYNTHETIC DATASETS

Implementation Details	Grid	Testing Scenario	Loss Function[1]	Peak Demand [%]	Peak Feed-in [%]	R^2 Score	Peak Demand Mismatch [h][2]	Peak Feed-in Mismatch [h][3]
Implementation 1: Hybrid Datasets **(27 Synchronous + 1 Real)**	C	High	0.067	1.603	1.912	0.969	-2	0
		Low	0.061	0.461	3.327	0.954	0	0
Implementation 1: Hybrid Datasets **(140 Synchronous + 1 Real)**	C	High	0.064	0.077	2.404	0.968	-3	169
		Low	0.048	3.635	0.164	0.967	0	0
Implementation 2: Hybrid Datasets **(27 Asynchronous + 1 Real)**	C	High	0.057	6.293	1.662	0.973	-3	169
		Low	0.105	0.834	14.466	0.904	0	0
Implementation 2: Hybrid Datasets **(140 Asynchronous + 1 Real)**	C	High	0.065	2.719	1.406	0.968	-3	169
		Low	0.055	0.392	2.314	0.963	0	0
Implementation 3: Synthetic Datasets **(27 Synchronous)**	C	High	0.030	1.889	4.568	0.990	-3	0
		Low	0.033	3.424	2.264	0.985	0	-170
Implementation 4: Synthetic Datasets **(140 Asynchronous)**	C	High	0.061	0.982	1.628	0.970	-3	169
		Low	0.053	2.356	6.361	0.961	0	0
Implementation 1: Hybrid Datasets **(27 Synchronous + 1 Real)**	D	High	0.068	3.370	1.528	0.968	-3	169
		Low	0.062	1.352	3.224	0.955	0	169

[1] Loss Function = weighted combination of MSE and MAE; MAE = mean absolute error; MSE = mean squared error [2] Peak Demand Mismatch = Predicted- Actual Peak Demand Timing (in Hours) [3] Peak Feed-in Mismatch = Predicted - Actual Peak Feed-in Timing (in Hours)

high predictive accuracy, with peak demand and feed-in errors below 3.5% across both testing scenarios, confirming the method's scalability with increasing system size.

The blending methodology also delivers substantial computational time savings. Generating 27 training scenarios via *urbs* required approximately 37.1 hours on an Intel® Xeon® Platinum 8360 HL CPU (96 cores, 3.0 GHz). In contrast, generating 27 synthetic datasets required only 234 seconds on the LRZ-HGX-A100-80×4 system (four NVIDIA A100 GPUs). Including the time to generate the single real dataset, the total hybrid dataset preparation time was 2.28 hours, representing a 93.9% time reduction compared to building the surrogate from scratch. Purely synthetic data approaches can completely eliminate this burden, if the accuracy trade-off is acceptable.

IV. CONCLUSION

This work extends our previous surrogate modeling framework for distribution grids by introducing a scalable approach to construct surrogate models for large DGs composed of multiple sub-grids. The method leverages synthetic ground truth from sub-grid surrogates, proving these can be merged to construct the main-grid surrogate instead of building it from scratch. Only a single set of aggregated input data from the large DG is needed to predict its net hourly demand. As shown by implementations 3 and 4, even in the complete absence of real ground-truth data, the ensemble reliably captures the peak demand and peak feed-in while preserving the temporal profile of the DG's net demand. The evaluation on the larger grid, grid D, confirms the scalability of the proposed approach and the time savings achievable compared to constructing the surrogate model from scratch. The current study is limited to uncoordinated grids with minimal energy exchange between sub-grids. In future work, we will extend the grid scalability to purely synthetic ground truth implementations and to coordinated grids with significant energy exchange between the sub-grids.

REFERENCES

[1] Federal Government of Germany, *Climate Change Act 2021: Intergenerational Contract for the Climate*, Report, Federal Government of Germany, Berlin, Germany, [Online]. Available: https://www.bundesregierung.de/breg-en/service/archive/climate-change-act-2021-1936846, 2021.

[2] Bundesnetzagentur, *Update: Verteilernetze bis 2045*, SMARD, Mar. 2025, [Online]. Available: https://www.smard.de/page/home/topic-article/444/215544/update-verteilernetze-bis-2045.

[3] B. Reveron Baecker and S. Candas, "Co-optimizing transmission and active distribution grids to assess demand-side flexibilities of a carbon-neutral german energy system," *Renewable and Sustainable Energy Reviews*, vol. 163, p. 112 422, 2022. DOI: https://doi.org/10.1016/j.rser.2022.112422.

[4] A. Mohapatra *et al.*, "*Surrogate framework for energy system modeling*," in *Proceedings of the 2025 IEEE Kiel PowerTech*, 2025. DOI: 10.1109/PowerTech59965.2025.11180532.

[5] G. Pjetri *et al.*, "*Graph neural network surrogates for energy system modeling*," in *IEEE ISGT Europe 2025*, Accepted for publication, 2025, pp. 1–6. DOI: 10.5281/zenodo.17358030.

[6] M. Pollitt *et al.*, *The Active Distribution System Operator (DSO) – an International Study*, Sep. 2022. [Online]. Available: https://cerre.eu/wp-content/uploads/2022/09/CERRE_DSO_FinalReport_2709.pdf.

[7] TUM-ENS, *urbs: A linear optimisation model for distributed energy systems*, TUM, 2019. [Online]. Available: https://github.com/tum-ens/urbs.

[8] S. Candas *et al.*, "Optimization-based framework for low-voltage grid reinforcement assessment under various levels of flexibility and coordination," *Applied Energy*, vol. 343, p. 121 147, 2023. DOI: 10.1016/j.apenergy.2023.121147.

[9] M. Hoffmann, L. Kotzur, and D. Stolten, "The pareto-optimal temporal aggregation of energy system models," *Applied Energy*, vol. 315, p. 119 029, 2022. DOI: 10.1016/j.apenergy.2022.119029.

[10] M. Wortsman *et al.*, *Model soups: averaging weights of multiple fine-tuned models improves accuracy without increasing inference time*, 2022. arXiv: 2203.05482 [cs.LG].

[11] H.-z. Wang *et al.*, "*Deep learning based ensemble approach for probabilistic wind power forecasting*," *Applied Energy*, vol. 188, pp. 56–70, 2017. DOI: 10.1016/j.apenergy.2016.11.111.