



Institut für Vernetzte
Energiesysteme



Characterizing Load Shedding Events in a Highly Renewable Power Grid: A Data-Driven Approach

A Master's thesis by
Friedmuth M. Kraus

Written at the
Institute of Networked Energy Systems
of the
German Aerospace Center (DLR)

Submitted to the
Institute for Chemistry and Biology of the Marine Environment
at the
Carl von Ossietzky University of Oldenburg

In Partial Fulfilment of the Requirements
for the Degree of
Master of Science

Supervisor: Dr. Bruno U. Schyska
Second Examiner: Prof. Dr. Carsten Agert

20th June, 2025

Contents

1. Introduction	1
2. Methods and Background	3
2.1. Considered Variables and definitions	3
2.2. The European Resource Adequacy Assessment (ERAA)	6
2.3. The scope of the predictive models	7
2.4. Evaluation metrics	8
2.4.1. Performance measures of binary classifiers	8
2.4.2. Error measures for numerical variables	9
2.4.3. Measures of the ranking of variables	11
2.4.4. Box-and-whisker plots	12
2.5. Established indicators of variable renewable energy shortages	13
2.6. Variable renewable energy droughts	14
2.6.1. The Constantly-Below-Threshold method	14
2.6.2. The Mean-Below-Threshold approaches	15
2.6.3. The Sequential Peak Algorithm	16
2.7. The considered VRE drought definitions	17
2.8. Positive residual load events	17
3. Data	18
3.1. The ERAA data	18
3.1.1. The Pan-European Market Modelling Database (PEMMDB)	19
3.1.2. The Pan-European Climate Database (PECD)	19
3.1.3. The ERAA 2022 dataset	19
3.1.4. The ERAA 2024 dataset	22
3.2. The power flow simulations	24
4. A novel method for the detection of electricity shortages	25
4.1. Linear support vector machines	25
4.2. Definition of the novel predictor of electricity shortages	28

5. Results and Discussion	29
5.1. The exact prediction of hours with electricity shortages	30
5.1.1. The exact prediction of hours with electricity shortages using variable renewable energy (VRE) droughts	30
5.1.2. The exact prediction of hours with electricity shortages using the residual load	33
5.1.3. Electricity shortage Prediction with the novel Indicator . . .	33
5.2. Prediction of Loss of Load Hours	34
5.3. Prediction of the most relevant weather years	36
5.4. Cross-validation of the data-driven approaches	38
5.5. Transferability to a different dataset	42
5.5.1. The exact prediction of hours with electricity shortages on the ERAA 2024 data	42
5.5.2. The prediction of annual LOLH on the ERAA 2024 dataset	45
5.5.3. The prediction of the most critical weather years on the ERAA 2024 data	47
5.6. Transferability to another dataset without renewed training	50
5.6.1. The exact prediction of hours with electricity shortages on the ERAA 2024 data with pretrained indicators	50
5.6.2. The prediction of annual LOLH on the ERAA 2024 data with pretrained indicators	51
5.6.3. The prediction of the most critical weather years on the ERAA 2024 data with pretrained indicators	52
6. Discussion	54
6.1. Discussion of the VRE drought indicators as predictors of electricity shortages	54
6.2. Discussion of the residual load approach as an indicator of electricity shortages	55
6.3. Discussion of the novel indicator	57

7. Exemplary application	64
7.1. The switch from reanalysis based data to climate projections in the ERAA 2024 and its impact on resource adequacy concerns	64
7.1.1. Analysis of the differences between the reanalysis and climate projection based datasets in the PECD 4.1	65
7.1.2. Analysis of the differences between temperature adjusted reanalysis data and climate projections in the PECD 4.1	70
7.2. Evaluation	73
8. Conclusion	74
Acronyms	77
Glossary	80
References	81
A. The European Resource Adequacy Assessment	93
A.1. Involved institutions	93
A.1.1. The European Network of Transmission System Operators for Electricity (ENTSO-E):	93
A.1.2. The European Union Agency for the Cooperation of Energy Regulators (ACER):	94
A.2. The legal foundations of the European Resource Adequacy Assessment	95
A.3. Methodology	96

1. Introduction

What would have gone unnoticed as just another foggy Wednesday in November a decade ago prompted several German news outlets to launch investigations on 6th November, 2024 (Janzing, 2024; Krapp, 2024; Rönsch, 2024; WDR, 2024). Not only were fog and clouds blocking sunlight from reaching photovoltaics (PV) panels, but conditions were also largely windless across the country (Kaspar et al., 2024). Germany was in the midst of an infamous so-called “Dunkelflaute” or dark doldrum. Solar and wind energy supplied just 3.1% of the electricity consumed on that day, compared to 42.0% over the year. As a consequence, day-ahead stock market prices briefly exceeded 800 €/MWh – more than ten times the average price of 78 €/MWh in 2024 (Burger, n.d.). In the end, the situation was managed without any real risk of a power outage, thanks to existing reserves and imports (Bundesnetzagentur, 2025). Nevertheless, a broad and sometimes heated public debate was triggered, most notably after similar dark doldrums occurred again in December and January (Klein, 2024; Saurugg, 2024). Is Germany prepared for such situations in a future where solar and wind energy are expected to provide a much higher share of the electricity demand, particularly if the electrification of the transport and heat sectors further increase this demand (Deutschlandfunk, 2025; Saurugg, 2024)? Had the country’s decision to phase out nuclear and eventually fossil energy sources been a mistake (Deutschlandfunk, 2025)? And what measures would be suitable to ensure a reliable and affordable power system in the years ahead (Deutschlandfunk, 2025; Rönsch, 2024; Saurugg, 2024)?

Fortunately, researchers and power system operators were already well aware of the issue. Since 2021, the European Network of Transmission System Operators for Electricity (ENTSO-E) annually conducts an elaborate European Resource Adequacy Assessment (ERAA) that analyses the risks of electricity shortages over a 10-year time horizon so that decision-makers can take informed action if the need arises (ENTSO-E, 2025b). For the validity of this assessment, it is essential to consider all relevant sources of uncertainty. In particular, the intrinsic variability of weather must be taken into account. Up until 2023, this was mainly done by evaluating weather data from around 35 years of reconstructed historical climate data, so-called reanalyses (ENTSO-E, 2025a, pp. 55–60). However,

as meteorologists have pointed out, historical weather scenarios may not realistically represent future weather conditions shaped by climate change (Kapica et al., 2024). Therefore, the latest European Resource Adequacy Assessment (ERAA) 2024 used weather data from projections based on global climate models (ENTSO-E, 2025c, pp. 64–66; M. Koivisto et al., 2023). These in turn come with their own uncertainties. All climate models have biases, so a multi-model ensemble should be used (H. C. Bloomfield et al., 2022). Additionally, they operate based on assumptions about future developments, most notably the Shared Socioeconomic Pathways (SSPs) defined for the Sixth Assessment Report of the United Nations’ Intergovernmental Panel on Climate Change (IPCC), from which scenarios of future greenhouse gas emissions and climate policies are derived (Intergovernmental Panel On Climate Change (Ipcc), 2023, pp. 12–14). Further variations arise from differing assumptions about generator characteristics and random outages of generators (ENTSO-E, 2025c, p. 46; M. J. Koivisto & Murcia Leon, 2022). In summary, future ERAAs are supposed to incorporate a wide range of weather years from an ensemble of climate models, each operating based on a set of SSPs, technical scenarios and random outages. This can easily add up to thousands of scenarios that need to be assessed. However, the detailed power flow simulations used in the main analysis of the ERAA are computationally so demanding that this volume of data can quickly overwhelm available resources (Biewald et al., 2025). Therefore, a fast method to detect electricity shortages can be highly beneficial. Beyond enabling research into such events without extensive computational resources, it could also support the preselection of particularly critical subsets of data for comprehensive analysis in the main part of a resource adequacy assessment.

In this thesis, I investigate five methods to detect electricity shortages without the need for elaborate power flow simulations. Their predictive performance is evaluated at three levels of accuracy, reflecting the specific requirements of different tasks: The ability to identify exact hours at which electricity shortages occur is relevant, if timing and patterns of such events are of interest. However, for assessing adequacy risks, it may suffice to know the aggregated duration of electricity shortages within a year. Finally, selecting the climate years with the highest adequacy risk requires only a proper ranking.

My analysis shows that established methods for detecting dark doldrums which

consider only the supply side of the electricity balance perform comparatively poorly (Kittel & Schill, 2024; Mockert et al., 2023). As electricity shortages occur exactly when electricity demand and supply are not balanced, it might not come as a surprise that an approach additionally incorporating demand data yields the best results (Otero et al., 2022). However, its reliance on demand data can limit its applicability, as such data is not always readily available. As an alternative, I developed a data-driven method that does not require demand estimates but instead uses temperature and time data as proxies. Provided sufficient training samples, the approach achieves a predictive performance close to that of the method dependent on load data.

In the following *Methods and Background* section, I explain the context in more detail and introduce the four investigated methods taken from other authors as well as the measures chosen for their evaluation. Section 3, gives an overview of the datasets used. In Section 4 I define my novel approach, which sets the stage for the main analysis. I begin by evaluating the predictive performance of the selected approaches across different datasets and tasks in Section 5 and finish by discussing their strengths and limitations in detail in Section 6. In Section 7, as an exemplary application of my approach, I assess the impact of the switch from reanalysis-based weather data to climate projections on adequacy concerns. Finally, Section 8 concludes the thesis.

2. Methods and Background

2.1. Considered Variables and definitions

Before describing the investigated methods, I will define some concepts that are used throughout this thesis.

Nameplate capacity: The nameplate capacity of an electricity generator is the maximal power output of a generator that is provided under benchmark conditions (U.S. Energy Information Administration (EIA), n.d.). For VRE sources like solar PV, it is often much higher than the typical power output the generator will provide (Mockert et al., 2023).

Capacity factors: The capacity factor, cf , of an electricity generator is the ratio of the available capacity to its nameplate capacity at a given point in time (“Capacity factor”, 2023). It can be generalized to a set of generators stating the ratio of their combined available capacity to their combined nameplate capacity. Given n generators with capacity factors cf_1, \dots, cf_n and nameplate capacities P_1, \dots, P_n their combined capacity factor is defined as

$$cf = \frac{\sum_{i=1}^n P_i cf_i}{\sum_{i=1}^n P_i}. \quad (1)$$

In this thesis, I mainly use the combined capacity factor of the major VRE sources, namely solar PV and on- and offshore wind. In the following, I will call it the combined capacity factor or cf .

Full load hours: The full load hours of an electricity generator are the hours the generator would have to produce at nameplate capacity to generate the total power it produces in a given time frame, often a calendar year (Heylen et al., 2018). One way to calculate them is by taking the sum of an hourly time series of capacity factors over a year. They can be aggregated analogously to capacity factors.

The residual load: The residual load of an energy system is often defined as the difference of electricity demand and supply by VRE sources (Schwab, 2015, pp. 22–23; Do et al., 2016). In this thesis the considered VRE sources for the calculation of the residual load are solar PV and on- and offshore wind power. Other definitions include further renewable energy sources such as hydro power and sometimes power plants for which it is assumed that they have to run, e.g., because they produce industrial process heat (Gerke, 2014).

Electricity shortages The situations studied in this thesis are situations during which electricity supply can not meet electricity demand at a certain node in the grid. This can be due to an overall lack of electricity production or transmission constraints in the grid. Several names for such situations are used in the literature, including ‘power supply-demand imbalance’, ‘energy shortage’ and ‘electricity supply shortfall’ (Biewald et al., 2025; North American Electric Reliability Corpor-

ation [NERC], 2024; Shen et al., 2024). I will use the term 'electricity shortage' as it is brief and more specific than "energy shortage". Other authors use the similar term 'energy shortfall' to refer to the residual load (van der Wiel et al., 2019).

Loss of load hours (LOLH): The loss of load hours (LOLH) of a power system are the expected number of hours in a year with an electricity shortage (Ibanez & Milligan, 2014). The term loss of load expectation (LOLE) is sometimes used interchangeably in the literature (Avdijaj et al., 2024). However, the LOLE can also refer to the number of days in a year in which electricity shortages occur (Ibanez & Milligan, 2014). To avoid confusion, I used the unambiguous term LOLH.

Capacity mechanisms: Capacity mechanisms are measures taken to ensure the availability of sufficient generation capacities to meet demand at all times, even if these capacities are not economically competitive. This is mostly done by subsidies to providers of generation capacities. (Simoglou & Biskas, 2023)

Climate Reanalyses: Climate reanalyses are consistent and complete datasets of climate variables that aim to recreate historical weather data and are based on observations. The need for reanalysis arises from the fact that the available measurements of climate variables are incomplete. While there is a lot of measurements available for some areas of the world, others are only sparsely sampled. This becomes especially relevant for data in the pre-satellite era. To obtain complete, gridded data sets from the available measurements, they are fed into modern numerical weather prediction models. The outputs of those models are then taken as recreations of the historical weather. It is, thus, important to keep in mind that reanalysis data is not raw recorded data, but has undergone a complex preparation step that introduces some uncertainties and biases. (Jeppesen, 2023)

Climate projections: Climate projections are predictions of weather realisations in the future that are mostly obtained from elaborate climate models. While modern climate models have remarkable complexity and can be trusted to achieve reliable results on the trends of the climate represented by statistical properties of

the weather, the chaotic nature of the Earth’s system itself renders it impossible to predict the weather with any certainty on the scale of years (Etling, 2008, p. 250; Slingo & Palmer, 2011). Therefore, the weather predicted by climate models for future years has to be interpreted as a possible realisation of the weather in the predicted climate.

To be usable for the assessment of resource adequacy, it is important that rare events are properly represented in those synthetic weather years as these can correspond to critical situations for power systems (H. C. Bloomfield et al., 2021). The rare events of interest here, such as long periods with low wind speeds, do not automatically coincide with those that climate modellers have in mind when they assess the representation of extreme weather events.

2.2. The European Resource Adequacy Assessment (ERAA)

The European Resource Adequacy Assessment (ERAA) is a thorough evaluation of the resource adequacy of the European power system with a time horizon of up to ten years (ENTSO-E, 2025b, p. 6). It is annually conducted by the European Network of Transmission System Operators for Electricity (ENTSO-E) as required by Regulation (EU) 2019/943 of the European Parliament and of the Council of 5 June 2019 on the internal market for electricity (recast) (2019). Since its first edition was published in 2021, its methodology was improved every year. It now consists of two main modules. First an economic viability assessment (EVA) is conducted that analyses which generation units are economically feasible in what time periods (ENTSO-E, 2025c, pp. 39–51). These results are then used as input for simulations of the predicted power system with different climate datasets and assumptions about technical parameters (ENTSO-E, 2025c, pp. 52–62). The main output are hourly sampled time series that indicate if the power demand could be served and if not how much of the load remained unmatched.

The reliability of the results is double checked by the European Union Agency for the Cooperation of Energy Regulators (ACER), a decentralized agency of the European Union (EU) which, unlike the ENTSO-E that in part works for the interests of transmission system operators (TSOs), has to work for the good of the

European people.¹ Interestingly, the ACER did not approve the first two editions of the ERAA, not only because they decided the model needed general improvement, but specifically because they claimed that risks were overestimated. This emphasises that adequacy assessments must neither over- nor underestimate the risks (ACER, 2022, p. 18; ACER, 2023, p. 19). While too optimistic results risk electricity shortages, too pessimistic ones might lead to the implementation of expensive, oversized capacity mechanisms. Actually, the ERAA is the official reference for adequacy concerns in the EU, which means that capacity mechanisms can only be implemented if the ERAA supports concerns that can not be resolved by other measures.² More details on the involved institutions, legal foundations and methodology of the ERAA can be found in Appendix A.

2.3. The scope of the predictive models

As described in the introduction, I studied methods that can predict electricity shortages from the input data of a resource adequacy assessment without needing extensive resources. Of course, this requires major simplifications. I decided to only look at methods, that do not use the network structure of the grid directly. Therefore, the approaches do not have an explicit spatial component. Instead they work with the input data, such as capacity factor time series, aggregated for a spatial area. I only looked at data, that is aggregated to the country level and all analyses in this thesis are done with data for Germany. So when I write about the combined capacity factor of solar and wind generators, I refer to the combined capacity factor of all wind and solar generators in Germany. For the aggregation I assume lossless and instantaneous transmission of electricity within the aggregated area ('Copperplate model').³

¹Regulation (EU) 2019/942 of the European Parliament and of the Council of 5 June 2019 establishing a European Union Agency for the Cooperation of Energy Regulators (recast), 2019, Article 1(3); ENTSO-E, n.d.-a.

²Regulation (EU) 2019/943 of the European Parliament and of the Council of 5 June 2019 on the internal market for electricity (recast), 2019, Articles 20 & 21.

³See the paragraph on capacity factors in Section 2.1 for the details of the combination of capacity factors.

2.4. Evaluation metrics

To analyse the predictive performance of the considered indicators of electricity shortages I used a range of different metrics that focus on three different levels of interest, reflecting the needs of different tasks. The levels are:

1. The prediction of exact hours at which electricity shortages occur.
2. The prediction of annual LOLH.
3. The skill at ranking scenarios according to their annual LOLH.

In all cases results of power flow simulations are used as reference (See Section 3.2 for more details on these simulations). More precisely, I looked at hours during which more than 1 MWh of electricity demand can not be served according to the power flow simulations. These are used as reference hours with an electricity shortage.

2.4.1. Performance measures of binary classifiers

To evaluate the performance of different methods to predict the exact hours with electricity shortages they were interpreted as binary classifiers that sample hours into hours with and without electricity shortages. A binary classifier that classifies samples as positive or negative can be evaluated using the number of samples it correctly classifies as positives (true positives (TP)), respectively negative (true negatives (TN)) and the number of samples it incorrectly classifies as positives (false positives (FP)), respectively negative (false negatives (FN)). Here, a positive sample is an hour with an electricity shortage. Three common performance metrics of binary classifiers are used in this thesis to quantify the performance of the methods: Sensitivity, precision and F_β -score.

Sensitivity: The sensitivity (or recall) of a binary classifier is the ratio of true positives to all positive samples (Tharwat, 2021).

$$\text{sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

Here: The ratio of hours with an electricity shortage that were detected to all hours with an electricity shortage.

Precision: The precision of a binary classifier is the ratio of the number of true positives to the number of all samples that are classified as positive (Tharwat, 2021).

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (3)$$

Here: The ratio of hours with an electricity shortage that are detected to all hours for which an electricity shortage is predicted.

F_β -score: The F_β -score combines sensitivity and precision in one measure. The parameter β allows to prioritize one over the other. While the F_1 -score is simply the harmonic mean of precision and sensitivity, values of β above 1 prioritize sensitivity and values below 1 put a higher weight on precision (Christen et al., 2023).

$$F_\beta = (1 + \beta^2) \frac{\text{precision} \cdot \text{sensitivity}}{(\beta^2 \cdot \text{precision}) + \text{sensitivity}} = \frac{(1 + \beta^2)\text{TP}}{(1 + \beta^2)\text{TP} + \beta^2\text{FN} + \text{FP}} \quad (4)$$

Here, the F_2 -score was used alongside the F_1 -score, reflecting the assumption that for certain tasks it is more important to detect most hours with an electricity shortage than to avoid predicting some where none occur.

2.4.2. Error measures for numerical variables

For the evaluation of the prediction of annual LOLH, I used established error measures for numerical variables. When predicting a numerical variable x_i with predictions \hat{x}_i , a wide range of error measures are available. An overview of common measures and their capabilities and deficiencies is given by Hyndman and Koehler (2006). The methods they introduce can be categorized into two types. For one, they introduce scale dependent measures, that change when a dataset is scaled and often have the dimension of the variable. While these measures are widely used they can not be used to directly compare the quality of predictions for different datasets. Secondly, they introduce a range of measures that adjust

the scale dependent ones to be independent of scale and allow for comparisons between different datasets. I will use the mean absolute error (MAE), an intuitive but scale dependent measure, and the mean absolute scaled error (MASE), a more complicated but scale independent measure.

Mean absolute error (MAE): The MAE is defined as

$$\text{MAE} = \text{mean}(|x - \hat{x}|) = \frac{1}{n} \sum_{i=1}^n |x_i - \hat{x}_i|, \quad (5)$$

where n is the number of samples (Hyndman & Koehler, 2006). It is recommended by Willmott and Matsuura (2005) as error measure for its straight forward interpretation as average absolute error, specifically compared with the very common root mean square error (RMSE). The MAE has the same dimension as x and depends on the scale of the data, which also facilitates its interpretation but prohibits comparisons of the predictive performance of algorithms on different datasets.

Mean absolute scaled error (MASE): The MASE adjusts the MAE to be independent of scale by dividing the errors with the MAE of a simple benchmark prediction method. The original paper suggests theso-called naïve forecast method as benchmark for time series data (Hyndman & Koehler, 2006). In my case I use it to compare weather years that are treated as examples of possible weather realizations that do not always constitute a time series as they might stem from different climate projections. Therefore, I use the version for general data samples with the arithmetic mean of x , \bar{x} , as benchmark prediction, as was proposed in a forum post by Hyndman (<https://stats.stackexchange.com/users/159/rob-hyndman>) (2014). The resulting definition of the MASE is:

$$\text{MASE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{x_i - \hat{x}_i}{\frac{1}{n} \sum_{j=1}^n |x_j - \bar{x}|} \right| = \frac{\frac{1}{n} \sum_{i=1}^n |x_i - \hat{x}_i|}{\frac{1}{n} \sum_{j=1}^n |x_j - \bar{x}|} = \frac{\text{MAE}}{\text{MAD}}, \quad (6)$$

where $\text{MAD} = \frac{1}{n} \sum_{j=1}^n |x_j - \bar{x}|$ is the mean absolute deviation. So, the MASE is simply the MAE scaled by the mean absolute deviation (MAD) in my case.⁴ Its values can be interpreted as showing how well the prediction is on average com-

pared to always predicting the mean of the measurements. Values greater than 1 indicate a worse prediction and values below 1 a better one. While the prediction of the mean seems to be trivial, it should be noted that it is possible that no available approach achieves a comparably well prediction, as it uses information on the measurements that will often not be available to other methods. Therefore, a MASE value above 1 does not necessarily indicate a bad predictive performance. The MASE is proposed as a standard measure for the evaluation of forecast accuracy by Hyndman and Koehler (2006). Unlike similar measures, such as the mean absolute percentage error (MAPE) or the mean relative absolute error (MRAE), it is always defined and finite except for trivial cases where the data is constant.

2.4.3. Measures of the ranking of variables

Sometimes, the absolute values of a predicted variable are not relevant but only the ranking assigned to a data point, so, the order of data points from the lowest to the highest value. This can for example be the case in the context of resource adequacy assessments when the most critical scenarios for a thorough analysis have to be selected and the number of scenarios that can be processed is already fixed. I use two measures in this context, one that assesses the overall similarity of the ranking given by a prediction and a reference and one that focuses on the similarity for a certain amount of the highest ranks.

The Spearman rank correlation: The Spearman rank correlation of two variables, X and Y is defined as the Pearson correlation coefficient of the rankings of the variables (Myers & Well, 2003, p. 508). If ties occur, all data points with the same value are assigned an average rank (Dodge, 2010, p. 502). Hence, if exactly n data points have a lower value than m other data points that have the exact

⁴This makes the MAE equivalent to the relative mean absolute error (RelMAE) with the mean as benchmark prediction in my case. The original context in which the MASE was developed is the prediction of time series, for which in many cases previous parts are available. In that context, the RelMAE uses the MAD computed on the previous data ("out of sample") while the MASE computes the MAD of the sample ("in sample").

same value, these m data points all get the rank

$$\frac{1}{m} \sum_{i=1}^m n + i.$$

Like the Pearson correlation coefficient, it varies between -1 and 1 , where 1 indicates a perfect match of the rankings and -1 an opposite ranking.

The Sørensen-Dice coefficient: The Sørensen-Dice coefficient or Dice similarity coefficient (DSC), originating in ecology, was developed as a measure of the similarity of two statistical samples (Dice, 1945). Given two finite samples A and B , the DSC is defined as:

$$DSC(A, B) = \frac{2|A \cap B|}{|A| + |B|} \in [0, 1]. \quad (7)$$

In the case of finite sets it is equal to 1 if and only if both sets are identical and equal to 0 if and only if the sets do not intersect. I use it to quantify the ability to detect the most critical weather years by computing the DSC of the two sets of the n weather years with the most LOLH predicted by the analysed indicator and the reference. Here, n can be varied according to the number of weather years that are considered to be of interest.

2.4.4. Box-and-whisker plots

Box-and-whisker plots are a very common way to depict the univariate distribution of a data set. Because their precise definition can vary between publications I will give a short explanation of the definition used in this thesis. As depicted in Figure 1, the box is delimited at its upper and lower end by the first and third quartile of the underlying empirical distribution. Between those two is a third line representing the median. The length of the box, so the difference between third and first quartile, is called inter quartile range. The whiskers of the box extend to the data point furthest away from the quartiles that are within a 1.5 inter quartile range distance from the quartiles. All data points outside this range are considered to be outliers and are depicted by small circles (du Toit et al., 1986, p. 29).

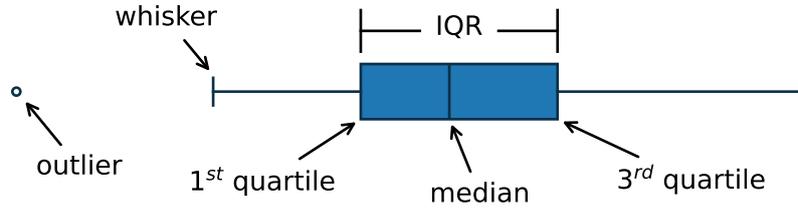


Figure 1: Illustration of a box-and-whisker plot. IQR stands for the inter quartile range.

2.5. Established indicators of variable renewable energy shortages

The challenges that VRE sources pose for a power system that heavily relies on them are subject of research for several decades now and over the years, several viewpoints and definitions of critical situations caused by the variability of VRE sources have been established. Kittel and Schill (2024) assembled a comprehensive collection of different types and measures of variable renewable energy shortages. I will follow the nomenclature proposed in their work where they use the term 'variable renewable energy shortage' or 'VRE shortage' as an umbrella term for different types of events during which VRE sources are scarce. They identified four subtypes of VRE shortages:

1. **Variable renewable energy droughts:** VRE droughts are focused on the supply side of the energy balance and only consider availability time series of VRE sources. They are defined using a threshold under which the availability time series has to fall. The considered time series can be relevant meteorological data such as wind speeds or further processed data such as capacity factors and while some definitions consider only one energy source others include a broader range.
2. **Positive residual load (PRL) events:** As their name gives away, positive residual load events additionally include the energy demand side and are focused on the balance of supply and demand. Situations where electrical load exceeds VRE supply are called PRL events.

3. **Variable renewable energy anomalies:** Focused on the supply side, VRE anomalies are defined by the accumulated deviation of VRE availability time series from a reference over a given time frame.
4. **Electricity system stress events:** Electricity system stress events take an economic point of view and are defined by high electricity prices, which are understood to indicate VRE shortages that create system stress.

VRE anomalies classify whole periods of time, so they are not well adapted to the task of identifying exact hours at which electricity shortages occur. Electricity system stress events on the other hand rely on market data, which is not provided as input to the ERAAs. Therefore, they can not be used for a pre-selection of relevant scenarios. Thus, I restricted my analysis to VRE droughts and PRL events in this thesis.

2.6. Variable renewable energy droughts

All VRE drought definitions given by Kittel and Schill (2024) have in common that they use a time series characterizing VRE supply and identify situations in which that time series falls below a given threshold $\theta \in [0, 1]$. I always use the combined capacity factor of on- and offshore wind and solar PV, cf , as time series for the VRE droughts. Furthermore, Kittel and Schill (2024) are not only interested in the individual hours of VRE droughts, but aggregate them to events defined by consecutive hours in which the drought condition holds. This can be important in the analysis of VRE droughts because longer droughts have the potential to become much more critical to the power system. In the following the different definitions of VRE droughts are introduced.

2.6.1. The Constantly-Below-Threshold method

A straightforward way to define a VRE drought is to consider every point in time where cf is below the threshold to be part of a VRE drought. Kittel and Schill (2024) call this a Constantly-Below-Threshold (CBT) VRE drought, as the capacity constantly has to stay below the threshold during the VRE drought.

They further point out though, that the CBT method tends to underestimate the size of energy droughts, as already one sample above the threshold ends the event. This can lead to the identification of a multitude of small VRE droughts shortly after one another in a situation where the relevant time series fluctuates around the threshold. Each one of these droughts might be harmless if seen as isolated events. Capacity reserves and storage solutions such as pumped hydro storage or batteries might be expected to be able to compensate for the lack of VRE sources for such short times. The accumulation of those events can, however, pose a much greater risk to system stability, as the short interim with higher energy availability might not suffice to refill energy storage. Thus, it is often more insightful to combine these events and consider them as one big event. This process, called pooling, can be done by various means, some of which will be introduced below.

2.6.2. The Mean-Below-Threshold approaches

One pooling procedure relies on a running mean to determine if VRE drought events should be pooled or considered to be separate events. The Mean-Below-Threshold (MBT) method, applies a running mean to the time series before comparing it to the threshold. The publications Kittel and Schill cite for this definition differ in their details. Some authors define the VRE drought as all points in time where the running mean falls below the threshold (Potisomporn et al., 2024). Another definition is given by Mockert et al. (2023). Every point in time that contributes to a running mean below the threshold is associated with a dunkelflaute by them.

A common approach to the MBT method is to choose a fixed time span T_{MA} for the running mean. Kittel And Schill refer to this as the fixed-duration mean below threshold (FMBT) approach. They criticize that the choice of T_{MA} is arbitrary and argue that the approach is of limited utility as the results for different T_{MA} vary significantly. To avoid this, they propose their own variable-duration mean below threshold (VMBT) approach. They start with a big initial T_{MA} and repeat the whole process for successively decreasing values of T_{MA} . This way, they claim to capture VRE drought events on all temporal scales without ambiguities that the choice of a parameter would create. This can only be guaranteed if they start

with a sufficiently large value of T_{MA} , such that no higher values of T_{MA} would find any VRE drought events. To achieve this, they claim that it suffices to start with any large value of T_{MA} such that no drought event is detected for it. However, this is not true in general as can be proven by counter example. Suppose the capacity factor were given by

$$cf(t_i) = \begin{cases} 0.01 & \text{if } i \text{ is divisible by } 20 \\ 0.2 & \text{otherwise.} \end{cases} \quad (8)$$

Given a threshold $\theta = 0.19$ and a running mean with length $T_{MA} = 20$, the running mean would be at a constant level of $(19 \times 0.2 + 0.01)/20 = 0.1905$, above the threshold. Setting $T_{MA} = 21$ the running mean would attain values of about 0.1819, below the threshold. This counterexample albeit a bit artificial shows that the VMBT approach is not as unambiguous as they assume. Furthermore it would be necessary to take every possible smaller value of T_{MA} to ensure, that no droughts are missed. This makes this approach quite costly if started with a really high value. I would conclude that the approach allows to capture droughts on different scales but does not provide the unambiguity Kittel and Schill (2024) promise without significantly increased use of resources. Therefore, I decided to use the FMBT approach used by Mockert et al. (2023) for my evaluation of the MBT pooling approach.

2.6.3. The Sequential Peak Algorithm

A different pooling approach is given by the Sequent Peak Algorithm (SPA), originating in hydrology. It utilizes the cumulative energy deficit of a drought event and does not need additional parameters. With the SPA method an event starts when the considered time series, e.g. cf , falls below the threshold. Afterwards a cumulative deficit function ED^{SPA} is calculated:

$$ED^{SPA}(t_i) = \max(0, ED^{SPA}(t_{i-1}) + \theta - cf(t_i)) \quad (9)$$

The VRE drought event is then defined to end when ED^{SPA} attains its maximal value before the next zero. Due to its independence of parameter choice beside the choice of a threshold θ , Kittel and Schill see the SPA as one of the best methods

to define VRE droughts, together with their own VMBT method.

2.7. The considered VRE drought definitions

In total I decided to consider three VRE drought definitions:

1. The FMBT method as defined by Mockert et al. (2023), with a running mean of 48 h and a threshold of 0.06.
2. A variation of the first, with a threshold of 0.12 and a running mean of 72 h. These parameters were chosen after some experimentation with the goal to increase the amount of electricity shortages that are correctly detected.
3. The SPA algorithm with a threshold of 0.06.

2.8. Positive residual load events

Another approach to the detection of critical situations in energy grids are positive residual load (PRL) events, i.e., situations in which electricity demand exceeds supply by VRE sources. As most energy systems nowadays still have a relevant share of non-renewable energy sources, the residual load is positive most of the time. So, actually, high residual load event might be a better term for the situations of interest.

A straightforward approach to define high residual load events is proposed by Otero et al. (2022). They define the days with the highest 10% of residual loads as energy droughts. Using the terminology of Kittel and Schill (2024) this is similar to a CBT method with a threshold set to the 0.9-quantile of the dataset. Of course in this case the approach would have to be changed to not identify the data points below but above the threshold. As the LOLH are far below 10% of the total time, this exact definition will systematically overestimate the resource adequacy concerns. Therefore, I consider the quantile at which to set the threshold to be a hyperparameter to optimize, similar to the approach taken by Biewald et al. (2025). This also addresses the criticism that the CBT approach can produce somewhat arbitrary results due to the need to choose a threshold (Kittel & Schill, 2024). To an extent this issue is resolved because a criterion for the choice of

the threshold is given. However, the solution is not perfect, as training data with reference values for hours with electricity shortages is required and the threshold depends on the particular dataset in question. To achieve comparability to the other considered approaches, I, furthermore, adjusted the method to use an hourly sampled time series instead of a daily sampled one.

By its definition this approach will always detect the same percentage of hours with electricity shortages. This can be detrimental to its prediction skill on other data sets than the one the quantile is optimized for or even subsets of the data set it is optimized for. Consider, for example, a subset of the original dataset consisting of a single weather year. The approach will always detect the same LOLH no matter if the year is a particularly critical one or not. To avoid this shortcoming, I adjust the approach by setting a threshold defined by the optimized quantile, so a certain level of the residual load, instead of using the percentage of critical data points as definition. I will refer to this definition as the residual load approach after Otero et al. (2022) or simply residual load approach in the remainder of this thesis.

3. Data

The data used in this thesis is mainly the publicly available input data of the European Resource Adequacy Assessments complemented by climate data from two other sources and results of economic dispatch (ED) power flow simulations conducted at the German Aerospace Center (DLR). While the power flow simulations provide reference predictions of electricity shortages, the ERAA input data supplies time series of capacity factors of VRE sources, temperature and electricity demand.

3.1. The ERAA data

On the highest level, the data is divided into two datasets: one compatible with the ERAA 2022 and one with the ERAA 2024. Before describing them in more detail I will briefly introduce two publicly available databases in which most of the input data of the ERAAs is published .

3.1.1. The Pan-European Market Modelling Database (PEMMDB)

The data on installed capacity resources and interconnection capacities is compiled by the TSOs based on the National Energy and Climate Plans (NECPs), European plans such as the "Fit for 55"-target of the EU as well as estimates from TSOs and distribution system operators (DSOs) and independent research institutions. It is then collected, harmonized and amended by the ENTSO-E and made available to the public in the so-called Pan-European Market Modelling Database (PEMMDB). (ACER, 2020, Article 5)

3.1.2. The Pan-European Climate Database (PECD)

The climate data used for the ERAA was compiled by the European Centre for Medium-Range Weather Forecasts (ECMWF) and is published in the Pan-European Climate Database (PECD) (Copernicus Climate Change Service (C3S) Climate Data Store (CDS), 2024). It includes several climate variables such as wind speeds on different altitudes, air temperature and solar surface irradiation that can be attained as spatially gridded data or aggregated to time series on bidding zone or (sub-)country level. All variables are available from different sources. At first the only source of meteorological data was the ERA5 reanalysis, created by the Copernicus Climate Change Service (C3S) at the ECMWF (Copernicus Climate Change Service, 2019). In version 4.1 of the PECD, meteorological data from three CMIP6 climate models are introduced in accordance with the official ERAA methodology (ACER, 2020, Article 4 (e)). The climate variables are further processed to attain hourly capacity factor time series aggregated to the available levels. The wind power availability time series were created at the Technical University of Denmark (DTU), while the solar PV capacity factors were estimated at the Paris Sciences et Lettres University (PSL) (M. Koivisto et al., 2023).

3.1.3. The ERAA 2022 dataset

The available network infrastructure is defined in the corresponding PEMMDB available at the website of the ENTSO-E (n.d.-b). The electricity demand data was also taken from this location. For my analysis, I focused on the target year 2030, so, the assumptions about the network infrastructure are predictions about

Resource Capacities [GW]	solar PV	onshore wind	offshore wind
ERAA 2022	215.0	110.0	30.2
ERAA 2024	97.0 + 118.0	115.0	28

Table 1: Installed VRE capacities in the PEMMDBs corresponding to the 2022 and 2024 ERAAs after the economic viability assessment (EVA). The given capacities pertain to Germany and the target year 2030.

the situation in 2030. Table 1 lists the most important figures from the PEMMDB for said target year in Germany.

The availability of the generation capacities is given by capacity factor time series compiled in weather years. Weather years refer to the year that the weather data is coming from. In case of the ERAA 2022 the climate data comes from the ERA5 reanalysis with a bias correction of the wind speeds based on the Global Wind Atlas in its second version, that was found to be more accurate than ERA5, especially in mountainous regions (Murcia et al., 2022). I used the 31 weather years from 1985 to 2015. This is supposed to account for the variability in weather data. Additionally, the DTU, that modelled the wind capacity factors, provided several variations of their results based on different assumptions about the distribution and technical specifications of the installed wind energy generators as part of the PECD 2021 update (M. J. Koivisto & Murcia Leon, 2022). The considered parameters of these scenarios are the hub height and specific power of the generators, as well as the resource grade of their locations. The resource grade describes the quality of the location, measured by the mean wind speed. The top 10% percent of locations in a bidding zone get resource grade A. Locations with mean wind speeds between the fiftieth and ninetieth percentile get resource grade B and the 50% of worst locations get resource grade C. I used 10 of the scenarios provided by the DTU as the results of power flow simulations for all 31 weather years were available to me:

The "existing" scenario: The "existing" scenario assumes that the distribution of the specific power, hub height and resource grade of the onshore wind generators are the same as in the year of the modelling (M. J. Koivisto & Murcia Leon, 2022). The product user guide of the PECD 4.1 states that its "existing" scenario relies on data pertaining to the year 2020 attained from the WindPowerNet (ECMWF, n.d.,

Section 2.9.1. “Wind energy database”, n.d.). As I could not find more specific information on the “existing” scenario of the PECD 2021 update, I assume it refers to the same data. Note, that this does not imply that the amount of installed capacities in the target year is the same as it were in 2020.

The uniform scenarios: Nine scenarios are defined by the assumption that all on-shore wind generators have the same specifications. These specifications are combinations of one of three hub heights, 100 m, 150 m and 200 m, with one of the three resource grades. The DTU additionally provides scenarios varying the specific power. I only had access to the results of power flow simulations of the scenarios with a specific power of 199 W/m^2 . The average specific power of onshore wind generators installed in Germany in 2024 was 302 W/m^2 and the overall fleet in Germany had a specific power of about 380 W/m^2 in 2020 with a negative trend coming from about 400 W/m^2 (Janal et al., 2025; Lüers & Heyken, 2024). So, the 9 scenarios assume a relatively low specific power and wind turbines with a lower specific power generally have higher capacity factors as they can operate more efficiently, particularly with lower wind speeds (Johansson et al., 2017). Therefore, the 9 scenarios will tend to overestimate the capacity factors of onshore wind power.

Additionally to the scenarios provided in the PECD, I used two other sources for capacity factor time series:

The Reading dataset: The Department of Meteorology of the University of Reading, Reading, UK, provides another dataset (H. Bloomfield & Brayshaw, 2021). It is also based on the ERA5 with a bias correction of the wind speeds to match the magnitudes of the Global Wind Atlas version 2 dataset (H. C. Bloomfield et al., 2022; “Global Wind Atlas”, n.d.).

The Renewables.Ninja dataset: Finally, capacity factors from the Renewables.Ninja website were used (Pfenninger & Staffell, 2016; Staffell & Pfenninger, 2016). Other than the previously described datasets, it is based on the MERRA-2 reanalysis provided by the NASA.

While scenarios for the offshore wind generation were available, I had only access to power flow simulations based on the reference scenario. Apart from that, I used the 2m air temperature provided by the ECMWF for the PECD 4.1 (Copernicus Climate Change Service (C3S) Climate Data Store (CDS), 2024).

3.1.4. The ERAA 2024 dataset

Again, the general assumptions about the network infrastructure are taken from the website of the ENTSO-E (n.d.-c). An important change in the ERAA 2024 data compared to earlier versions is hidden in the definition of the electricity demand data. Unlike in earlier versions of the data, times with negative loads occur. Sometimes these negative loads go up to about 20 GWh and while the average load before was at about 88 GWh it is now at only approximately 63 GWh. This sudden change is due to a decision to model all electricity storage units that are not participating in the electricity market as part of the demand. More precisely, this means that for example behind-meter batteries are accounted for as negative loads and not as storage units (ENTSO-E, 2024a, p. 27). This can be beneficial for TSOs as they can concentrate on modelling the units they can control. For other applications this new definition of the demand can be challenging, however. Unfortunately, the data provided is somewhat ambiguous. For instance, it was not clear at the time of the creation of the power flow simulations, if the 118.0 GW of rooftop PV capacities are also modelled as part of the demand. As the differences in the average electricity demand was so significant, it was assumed that this was the case. So, the power flow simulations assume only 97.0 GW of PV capacities in the market.

The meteorological data is now completely taken from the PECD 4.1, the first PECD to provide data from climate projections in addition to the reanalysis data (Copernicus Climate Change Service (C3S) Climate Data Store (CDS), 2024). I used data from all three available climate models: The CMCC-CM2-SR5 (CMR5) model, the EC-EARTH3 (ECE3) model and the MPI-ESM1-2-HR (MEHR) model (Döscher et al., 2022; Lovato & Peano, 2020; von Storch et al., 2017). Again, several technical scenarios are provided, this time given id numbers:

scenario id	31	32	33	34	35	36	37	38	39
specific power [W m^{-2}]	199	199	199	277	277	277	335	335	335
hub height [m]	100	150	200	100	150	200	100	150	200

Table 2: Technical specifications of the uniform onshore wind generation scenarios in the PECD 4.1.

scenario id	21	22
specific power [W m^{-2}]	316	370
hub height [m]	155	155

Table 3: Technical specifications of the uniform offshore wind generation scenarios in the PECD 4.1.

The "existing" scenarios: As stated before, the "existing" scenario assumes that the distribution of the technical parameters of the wind generators is the same in the target year as in 2020 (ECMWF, n.d., Section 2.9.1.). The id number of the "existing" scenarios for on- and offshore wind are 30 and 20, respectively.

The uniform scenarios: Again, these scenarios assume all on- or offshore wind generators to have the same technical specifications. Other than in the PECD 2021 update, the resource grade is not varied but always set to B (ECMWF, n.d., Section 2.9.1.3.). The specifications for specific power and hub height of the 9 onshore and 2 offshore wind scenarios are shown in Table 2 and Table 3, respectively.

While meteorological data for the 51 weather years from 2015 to 2065 are available for the climate models, I had only access to time series for the electricity demand for the 12 weather years from 2025 to 2036 that were provided to the DLR by a TSO. Therefore, the main part of my analysis uses only these 12 weather years and only the "existing" scenarios for the technical specifications, as they were published first. However, for the exemplary application of my indicator in section 7, I studied all 30 combinations of technical on- and offshore wind scenarios for all 42 years of historical data and the 50 years from 2015 to 2064 of the projections.

3.2. The power flow simulations

To assess the quality of the predictions of the considered indicators, a reference for the occurrence of electricity shortages is necessary. While the ERAA provides energy not served (ENS) time series, they are not available for the various technical scenarios I wanted to analyse. Instead, I used results of a power flow simulation model of the European power grid that was originally developed as part of the Destination Earth (DestinE) project of the EU by the DLR (Schyska, Bruno U. et al., 2024).

The model is based on the PyPSA package of the Python programming language, that provides a comprehensive open-source environment for the modelling of power systems (Brown et al., 2018). Compared to the models employed by the ENTSO-E, the model is heavily simplified and aggregates all variables to the country level. So, while the transmission of electricity between countries is explicitly modelled and constrained by the available network capacities, the transportation of electricity within countries is assumed to be instantaneous and unconstrained ("copper plate model"). The optimisation of the model is done for a whole year as an economic dispatch with a linear optimal power flow and perfect foresight. So, the result will use the available resources in a way that minimizes the total costs over the year and provides the demanded energy supply whenever possible while modelling the power flow linearly. Generation units are assumed to have no ramping constraints, i.e., they can be switched on and off immediately. Despite these simplifications the results are quite similar to the ones obtained by the TSOs if these are aggregated to the country level.

While the simulations for the ERAA 2022 dataset were conducted with the original model from the DestinE project, the computations for the ERAA 2024 dataset were done with an updated model that was benchmarked to achieve results as close as possible to ones obtained by a TSO as part of a joint project. I did not conduct any of these simulations myself, but only used the results.

4. A novel method for the detection of electricity shortages

The residual load approach has one major caveat. It depends on demand data, that is not always available. On the other hand, my first analyses using VRE drought definitions clearly suggested that the demand side has a significant influence on the occurrence of electricity shortages that can not be captured by only considering indicators of supply. Therefore, I developed a data-driven novel predictor of energy shortages that uses readily available temperature and time data as proxies for the electrical load. This way, it incorporates the demand side of the electricity balance without the need for explicit load time series. This has been helpful for example for the evaluation of the PECD 4.1 climate data, as demand predictions for the dataset were not publicly available in June 2025, to the best of my knowledge.

As part of my algorithm, I needed a method for the classification of samples. I tried different established methods and decided to use linear support vector machines (SVMs), as they show one of the best results while being rather simple and therefore resilient to overfitting. Before explaining the details of my novel method, I will give a brief introduction to linear SVMs following the work of Zhou (2021, pp. 130–142).

4.1. Linear support vector machines

Given a set of data samples $x_1, \dots, x_n \in \mathbb{R}^m$ with labels $y_1, \dots, y_n \in \{-1, 1\}$, the aim of a linear support vector machine is to find a hyperplane that separates samples with negative and positive labels as well as possible. Like all learning algorithms, it has an inductive bias that, here, defines what exactly is meant by a good separation (Zhou, 2021, pp. 7–11). In the case of a training set that can be separated, SVMs will maximise the distance of the hyperplane to the closest training points. To be able to put that into a formula, at first, the hyperplane can be defined as all points $x \in \mathbb{R}^m$ that solve an equation of the form

$$w^\top x + b = 0 \tag{10}$$

where $w \in \mathbb{R}^m \setminus \{0\}$ is the normal vector of the hyperplane, which controls its orientation, and $b \in \mathbb{R}$ is its distance to the origin. The distance $d(x)$ of any point $x \in \mathbb{R}^n$ to the plane is given by

$$d(x) = \frac{|w^\top x + b|}{\|w\|}. \quad (11)$$

The condition that the samples with negative labels are on one side of the hyperplane and the samples with positive labels on the other can then be formulated as

$$\begin{cases} w^\top x_i + b > 0, & \text{for } y_i = 1, \\ w^\top x_i + b < 0, & \text{for } y_i = -1. \end{cases} \quad (12)$$

or equivalently

$$y_i(w^\top x_i + b) > 0, \quad \forall i. \quad (13)$$

The objective for the training of the SVM is then to find $w \in \mathbb{R}^m \setminus \{0\}$ and $b \in \mathbb{R}$ that maximise the minimal distance to any of the training samples while respecting Condition (13):

$$\begin{aligned} & \max_{w,b} \min_i d(x_i) \\ & \text{such that } y_i(w^\top x_i + b) > 0, \quad \forall i. \end{aligned} \quad (14)$$

This can be further simplified by using that the solution to Problem (14) is not unique. Actually, if $w \in \mathbb{R}^m \setminus \{0\}$ and $b \in \mathbb{R}$ define a hyperplane, then $c \cdot w$ and $c \cdot b$ define the same hyperplane for all non-zero $c \in \mathbb{R}$, as can be seen by multiplying Equation (10) with c . Now, if $w'_0 \in \mathbb{R}^m \setminus \{0\}$ and $b'_0 \in \mathbb{R}$ are a solution to Problem (14) and d_o the distance of the closest training point to the hyperplane, we can always replace w'_0 and b'_0 by $w_0 = d_o^{-1} \|w'_0\|^{-1} w'_0$ and $b_0 = d_o^{-1} \|w'_0\|^{-1} b'_0$. Using Equation (11), we get that the assumption

$$d(x_i) > d_o, \quad \forall i \quad (15)$$

is equivalent to

$$|w_0^\top x_i + b| > d_o \|w_0\| = d_o d_o^{-1} \|w'_0\|^{-1} \|w'_0\| = 1, \quad \forall i. \quad (16)$$

This means we can without loss of generality replace Condition (13) by

$$y_i(w^\top x_i + b) \geq 1, \quad \forall i, \quad (17)$$

with an equality for the samples closest to the hyperplane. The minimum in Problem (14) can then be rewritten as

$$\min_i d(x_i) = \min_i \frac{|w^\top x_i + b|}{\|w\|} = \frac{1}{\|w\|}. \quad (18)$$

So, we can replace Problem (14) with

$$\begin{aligned} & \max_{w,b} \frac{1}{\|w\|} \\ & \text{such that } y_i(w^\top x_i + b) \geq 1, \quad \forall i. \end{aligned} \quad (19)$$

As the norm is not differentiable and problems are often defined as minimisations, this is often replaced by the equivalent formulation

$$\begin{aligned} & \min_{w,b} \frac{1}{2} \|w\|^2 \\ & \text{such that } y_i(w^\top x_i + b) \geq 1, \quad \forall i. \end{aligned} \quad (20)$$

So far we assumed that the training data actually is separable by a hyperplane. As this is often not the case, Condition (17) is softened in the sense that it does not have to hold. Instead, its violation is factored in by a penalty in the optimisation. Different options for a penalty function are available. Notably the hinge loss, defined as

$$\ell_{hinge}(z) = \max(0, 1 - z), \quad \forall z \in \mathbb{R}, \quad (21)$$

which has nice mathematical properties for optimisations such as being convex. The modified objective function with hinge loss is then given as

$$\min_{w,b} \frac{1}{2} \|w\|^2 + \mathcal{C} \sum_{i=1}^n \max(0, 1 - y_i(w^\top x_i + b)), \quad (22)$$

where $\mathcal{C} > 0$ is a regularising constant that determines how strong the penalty for

violating Condition (17) should be. This problem can be solved very efficiently by optimised algorithms (Zhou, 2021, pp. 133–134).

The straightforward way to predict a class would now classify all samples on one side of the hyperplane as one class and all samples on the other side as the other, i. e., 1 if $w^\top x_i + b > 0$ and -1 otherwise. However, it can often make sense to use another threshold $\theta \in \mathbb{R}$ instead of 0. So, predict 1 if $w^\top x_i + b > \theta$ and -1 otherwise. This can, for example, be the case if the amount of samples of each class in the training set is imbalanced. Since the standard algorithm penalises each misclassified sample the same, it will tend to have a low sensitivity on the smaller dataset as each of the misclassified samples has a higher impact on the sensitivity. If this is not desired, the threshold can be tuned to optimise a scoring function, such as the F_β -score on a given set of training data. (He, 2013, p. 72; “Tuning the decision threshold for class prediction”, n.d.)

4.2. Definition of the novel predictor of electricity shortages

Using linear SVMs, I can now define the details of my method. As stated above, I used proxies of the electricity demand to incorporate it into my indicator. I tried to choose proxies that have a high impact on the variability of the demand and are readily available. In the end, I used the air temperature 2 m above the ground, the hour of the day and the information whether the day in question is a regular workday or not. Workdays are taken to be Monday to Friday, unless they are national German holidays. With an increasing electrification of heating systems, heat demand and therefore air temperature will become an increasingly relevant driver of electricity demand in the near future in Germany and similar countries with cold winters (Connolly, 2017). The daily routines of people and society have a relevant impact on demand as well (Castillo et al., 2022).

To capture the relationship between these proxies of the electricity demand with electricity shortages, I applied a data driven approach. Using the data on electricity shortages from scenarios and weather years that were simulated with the power flow model described in Section 3.2, an SVM classifier was trained. More precisely, the available training data is split into regular workdays and other days at first. These two subsets are again split by the hour of the day. Then, for each of the

48 resulting subsets, a linear SVM is trained to classify samples into hours with and without electricity shortages. The 2-dimensional space the SVM is trained on is spanned by the combined capacity factor in the first dimension and the 2m air temperature in the second.

I used the 'LinearSVC' class provided by the 'scikit-learn' package of the Python programming language as implementation of SVM classifiers (Pedregosa et al., 2011). As parameters I chose the standard values of the package: A regularisation constant $\mathcal{C} = 1$ and the square of the hinge loss as loss function for misclassified samples. Since my training data was heavily imbalanced with far less samples with electricity shortages than without, I had to tune the decision threshold of the classifier. Here, I used the F_β -score with $\beta \in \{1, 2\}$ as scoring function. Additionally, I standardised the temperature and the capacity factor in a preprocessing step, i. e., I subtracted the mean of the dataset from each value and divided it by the standard deviation. Thus, both features have a mean of 0 and a standard deviation of 1. This is a very common step in machine learning, ensuring that all features have the same impact on the optimisation (Han et al., 2012, pp. 113–115). Otherwise, the optimisation of the SVM would be as sensitive to an error of 0.1°C in the temperature dimension as an error of 0.1 in the dimension of the capacity factor.

To enable the SVMs to be trained, sufficient samples with energy shortages need to be in each of the 48 subsets of the training data. Subsets for which that is not the case are skipped until all other subsets are processed. Then, each subset that no SVM is trained for is assigned the SVM from the pool of trained ones that achieves the highest value of the scoring function on the training set.

5. Results and Discussion

Having introduced several candidates for the fast prediction of electricity shortages in section 2, the question of their predictive performance in different situations remains and is answered in this section. The level of accuracy that is of relevance here can differ between applications. For some applications, the exact hours at which electricity shortages occur are relevant. For others, the annual LOLH or the capability to select the most severe scenarios suffice. Thus, I assess different

metrics corresponding to these requirements.

As two of the approaches use electricity shortage data from power flow simulations for the optimization of their parameters, a cross-validation was conducted to analyse the dependence on the choice of training data. Furthermore, the analysis originally conducted on the data from the ERAA 2022 was repeated on the dataset from the ERAA 2024, with different weather years, technical scenarios and assumptions about the underlying power system. Finally, the models trained on the ERAA 2022 dataset are applied on the ERAA 2024 dataset to examine how well they can be applied if no training data is available.

5.1. The exact prediction of hours with electricity shortages

I evaluated the predictions of the exact hours with an electricity shortage by analysing the sensitivity, precision, F1- and F2-score of the considered indicators as predictors of electricity shortages obtained from power flow simulations. More precisely, I calculated these four metrics for the 31 weather years and 12 technical scenarios from the ERAA 2022 for which I had access to predictions from power flow based simulations as reference. The results of these calculations are depicted in the box-and-whisker plots in Figure 2. Each of the boxes shows the distribution of one of the metrics for one indicator over the 12 scenarios.

5.1.1. The exact prediction of hours with electricity shortages using VRE droughts

The first three indicators are the VRE drought or dunkelflaute definitions. The FMBT definition developed by Mockert et al. (2023) achieves sensitivities between 0.118 and 0.355 for the different scenarios, a precision ranging from 0.167 to 0.281 and F_2 -scores between 0.126 and 0.335. While the indicator is balanced between sensitivity and precision its overall performance is rather poor for this purpose. The adjustment of the FMBT approach with a longer running mean and higher threshold achieves better sensitivities between 0.566 and 0.891 but at the price of a high share of false positives resulting in a low precision between 0.076 and 0.118. The resulting F_2 -scores are comparable in the end, ranging from 0.254 to 0.385. Finally, the SPA, one of the two VRE drought definitions recommended

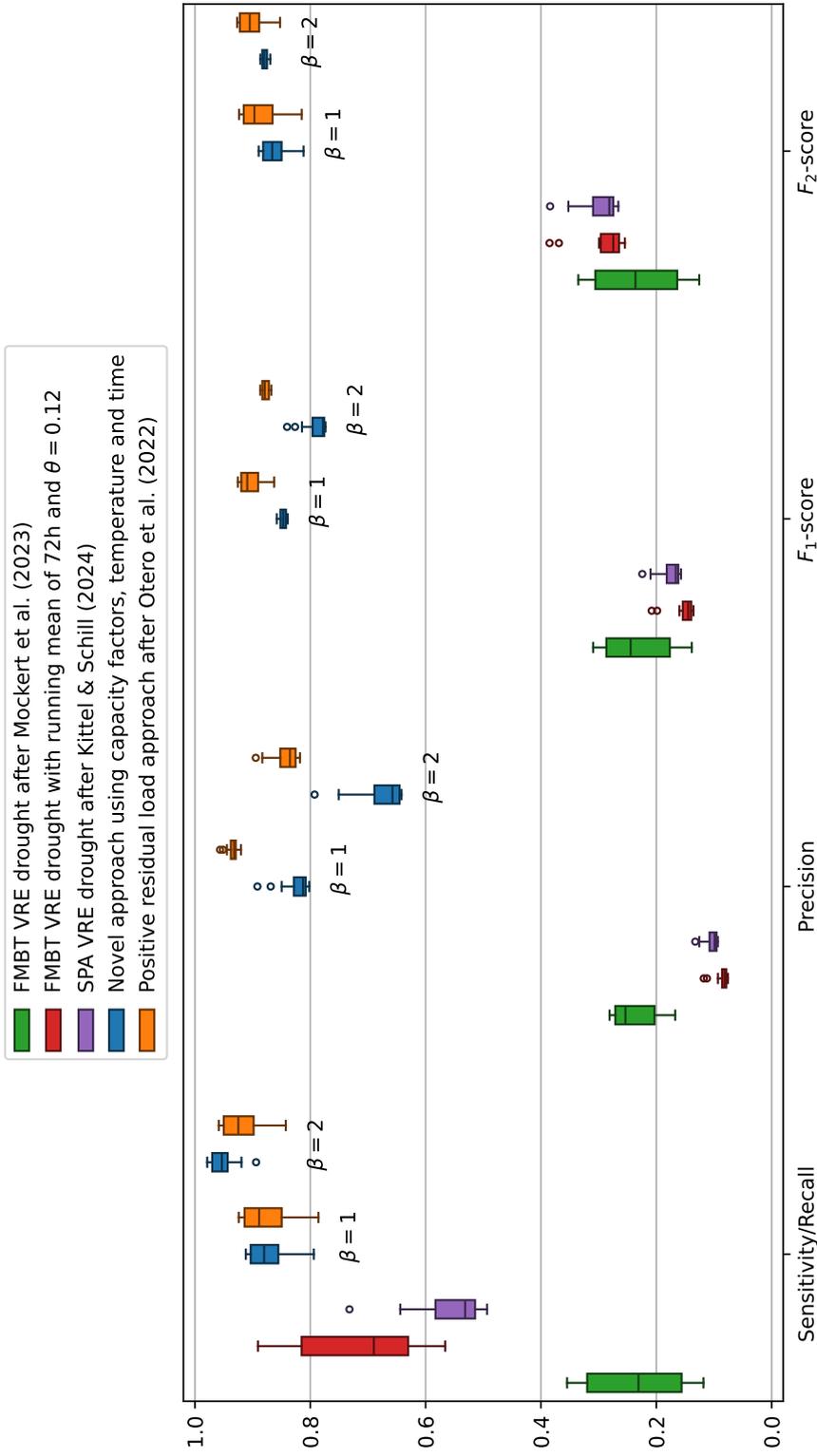


Figure 2: Box-whisker plots of the distributions of sensitivity, precision, F_1 - and F_2 -score of the five considered indicators. The 2 trained indicators are optimised for the F_β -score with two choices of beta: $\beta = 1$ and $\beta = 2$. The boxes show the distribution of the achieved values over the 12 scenarios in the ERA5 dataset.

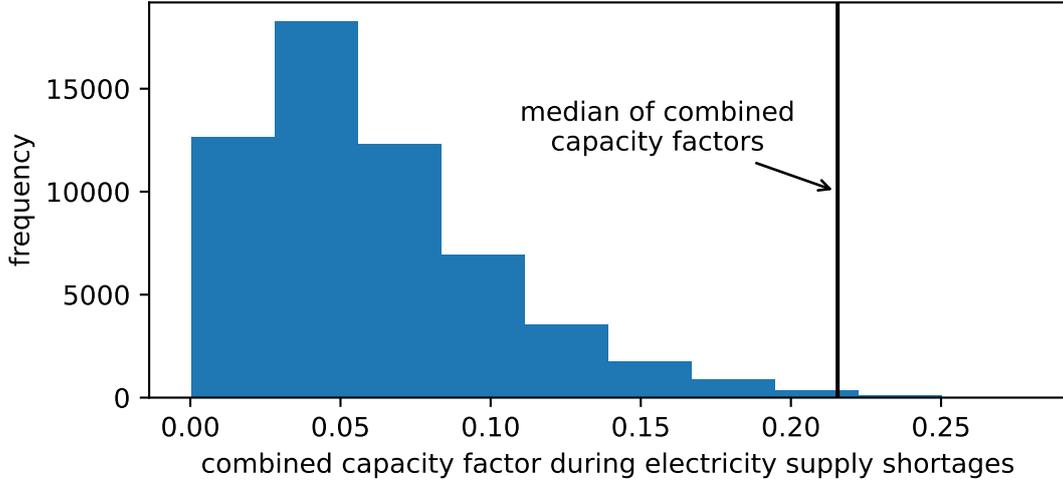


Figure 3: Histogram of the combined capacity factor for hours with electricity shortages in the ERAA 2022 dataset. The median of the general distribution of the combined capacity factor is indicated by a vertical line.

by Kittel and Schill (2024), gives similar results with a sensitivity between 0.494 and 0.732, a precision between 0.093 and 0.132 and thus F_2 -scores between 0.266 and 0.384. Overall, VRE drought definitions, that only consider the supply side of the electricity balance, are either too restrictive and miss most of the electricity shortages or greatly overestimate the total amount of LOLH.

Looking at the distribution of the combined capacity factor during hours with electricity shortages, depicted in the histogram in Figure 3, this does not come as a surprise. Electricity shortages do not only occur during times with extremely low capacity factors but even at capacity factors above the median of the distribution of the combined capacity factor at about 0.216. A threshold that would capture most of the loss of load hours would therefore detect more than 50% of the data points, while electricity shortages only occur for approximately 1.7% of the total time in this dataset. As mentioned earlier, it is therefore inevitable to include the second side of the electricity balance into the analysis: the demand.

5.1.2. The exact prediction of hours with electricity shortages using the residual load

The adjusted approach of Otero et al. (2022) described in subsection 2.8 incorporates the demand directly in form of the residual load. The optimization of the quantile above which residual loads are indicating an electricity shortage yields a residual load of about 104.3 GWh or 101.9 GWh as threshold for an optimization of the F_1 - or F_2 -score, respectively. The orange boxes in Figure 2 show the performance of the indicator with these thresholds. When the F_2 -score is optimized, the sensitivity has a higher weight than the precision. In this case, the sensitivity lies between 0.842 and 0.959, the precision ranges from 0.818 to 0.894 resulting in an F_2 -score between 0.852 and 0.927. The results for the optimisation of the F_1 -score are slightly worse for the sensitivity but better for the precision. In the end the F_β -scores are quite similar. This indicates an ability to adapt the approach to tasks where either of the metrics is to be prioritized over the other without compromising on the overall predictive performance too much. Overall, the approach achieves the best F_β -scores of all indicators.

A potential disadvantage of all approaches that use the residual load is their reliance on demand data as the estimations of the electricity demand are not always available. For example, the PECD 4.1 was updated in January 2025 including meteorological time series for 51 weather years (Copernicus Climate Change Service (C3S) Climate Data Store (CDS), 2024). However, in May 2025, the corresponding demand time series were still only publicly available for 12 of these years, to the best of my knowledge. To be able to analyse the full set of available weather years, an approach that indirectly incorporates the electricity demand without using the detailed demand data is needed.

5.1.3. Electricity shortage Prediction with the novel Indicator

My indicator, once trained, only uses the air temperature and time data as proxies of the electricity demand. Optimized to achieve the highest possible F_2 -score its sensitivity is better than that of any of the other considered indicators. Its precision is far better than the one achieved by the VRE drought indicators, but it is also significantly worse than the one of the residual load approach by Otero et al.

(2022). The F_2 -score is on average slightly lower than the one accomplished by the residual load method. However, due to the small range of the distribution over the 12 scenarios, the predictive performance for some of the scenarios is actually better. Optimizing the F_1 -score instead leads to greater changes than for the residual load approach. The precision increases by more than 10 percent while the sensitivity drops about 5 percent such that the resulting F_β -scores are, again, similar. This suggests a stronger ability to adapt to situations where one metric takes precedence over the other.

5.2. Prediction of Loss of Load Hours

In many situations the exact hours in which electricity shortages occur are not relevant. For the assessment of resource adequacy concerns, it suffices to know if electricity shortages occur and if so how many. The ENTSO-E, for example, focuses its analysis on the annual LOLH in its ERAA reports.

Figure 4 shows box plots of the MAE and MASE of the predicted annual LOLH of the five considered indicators. Again, the boxes show the distribution over the 12 technical scenarios of the ERAA 2022. Interestingly, the VRE drought definition after Mockert et al. (2023) achieves a median MAE of about 69.7 h and, thus, performs much better than the other two VRE drought definitions with median MAE values of approximately 477.7 h for the SPA and 850.8 h for the longer FMBT approach. An explanation for this is likely given by the comparatively well balanced precision and sensitivity of Mockert’s approach compared to the two other methods, that prioritise the sensitivity at the cost of an overestimation of the LOLH. This indicates that a good balance between sensitivity and precision might be beneficial for the task.

Looking at the MAE of my indicator and the residual load approach after Otero et al. (2022), it is evident that the approach using the residual load performs best, while my approach still achieves good results. I, again, optimized both indicators to maximize the F_1 - and F_2 -score and both indicators perform significantly better when optimized for the F_1 -score, which also supports the idea, that a balance between sensitivity and precision should be aimed for in this context. My indicator improves its median MAE by about 63% when the F_1 -score is optimised instead of

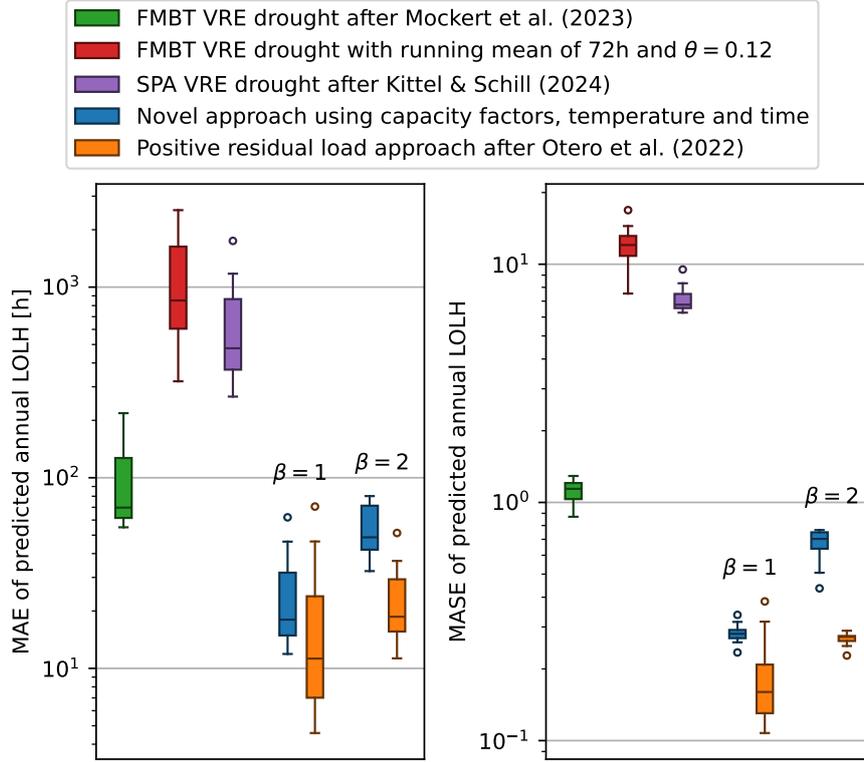


Figure 4: Box plots of the MAE and MASE of the predicted annual LOLH of the five considered indicators compared to reference results from power flow simulations for the ERAA 2022 data. The boxes show the distribution over the three climate models of the ERAA 2024 data.

the F_2 -score, while the residual load approach only improves by about 40%, again suggesting that my approach adjusts more flexibly to different optimisation goals.

The distributions of the MASE depicted in the right panel of Figure 4 are far more concentrated than the ones for the MAE. As is describe in section 2.4.2, in my case the MASE is simply the MAE divided by the MAD, a measure of the dispersion of the data. So, the concentrated distributions of the MASE show that the predictions of the indicators achieve a similar predictive performance for all scenarios relative to the variability of the data of the scenarios. More than that, the MASE can provide an intuition of the absolute predictive performance of the indicators as it compares the achieved MAE to the MAE of the method that always predicts the mean of the annual LOLH of the power flow simulations.

The FMBT approach after Mockert et al. (2023) has a median MASE of 1.14 and is therefore about as good as the mean method, while the other VRE drought definitions perform about an order of magnitude worse. The two indicators that incorporate the demand side of the electricity balance both achieve clearly better values if they are trained for the F_1 -score. In this case their MASE values are well below 1, indicating that they are able to predict a good part of the variability of the LOLHs.

5.3. Prediction of the most relevant weather years

For other tasks like the preselection of the most relevant scenarios for a more thorough analysis, it can suffice to predict a ranking of the relevance of the scenarios. In the context of resource adequacy assessments, I decided to take the annual LOLH of a scenario as measure of its relevance. I used two metrics to assess the ability of the considered indicators to detect the most relevant scenarios: The Spearman rank correlation of the annual LOLH predicted by an indicator with the reference assesses how well the indicator predicts the overall ranking of the scenarios. Additionally, the Sørensen-Dice coefficient of the sets of the ten scenarios with the most LOLH predicted by an indicator and the reference shows how well the indicators predict the ten most relevant scenarios. Figure 5 shows box-and-whisker plots of the distribution of the two measures over the 12 scenarios for which I had access to the results of power flow simulations.

Looking at the three VRE drought definitions first, their overall ranking performance is rather insufficient with median values for the Spearman rank correlation coefficients between 0.228 and 0.506. Interestingly, the definition after Mockert et al. (2023) clearly performs worst with values of ρ going as low as 0.024, although it performed best at the seemingly similar task of predicting the annual LOLH. This can be understood by looking at the median of the predicted annual LOLH of the other two definitions which is so much higher than that of the reference that the MAE is dominated by that bias. Apparently, they are, however, better at predicting the relative year-to-year variability of the variable, while Mockert's approach predicts the overall level of the annual LOLH well, but seems to fail at predicting its variability. Looking at the Sørensen-Dice coefficients, the approach

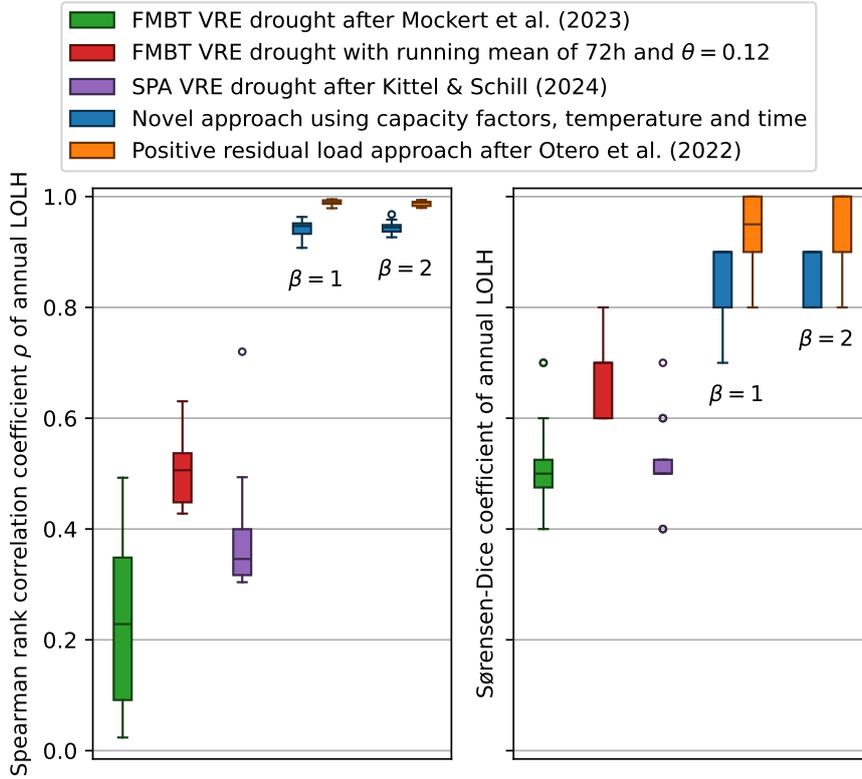


Figure 5: Box plots of the Spearman rank correlation coefficient (left panel) and the Sørensen-Dice coefficient (right panel) of the five considered indicators on the ERAA 2022 dataset. The boxes show the distribution of the achieved values over the 12 scenarios in the ERAA 2022 dataset. Each indicator was optimised for the F_β -score with two choices of beta: $\beta = 1$ and $\beta = 2$.

by Mockert et al. (2023) performs comparable to the SPA. Both achieve a median value of 0.5 indicating that 5 of the 10 weather years with the most annual LOLH of the reference were in the top ten weather years predicted by the indicators. So, Mockert’s approach seems to be better at predicting the most relevant scenarios than the overall ranking. The best Sørensen-Dice coefficient of the VRE drought definitions is achieved by the long FMBT approach, with a median value of 0.6. When looking at how many of the top 5 years of the reference are in the top 10 of the three indicators, they only detect 3 to 4 in the median.

The two approaches that also include the demand side perform much better with median Spearman rank correlation values between 0.945 and 0.990. Again, the

approach after Otero et al. (2022), that directly uses the residual load performs best and achieves values above 0.979 for all scenarios. The choice of β for the optimization of the F_β score does not have a big effect on the resulting values of Spearman’s ρ . The median is slightly better for $\beta = 1$ but the minimum is higher for $\beta = 2$. For the Sørensen-Dice coefficient, my approach achieves a median of 0.9 for both values of β , while the median for the residual load approach is 1 for $\beta = 2$ and 0.95 for $\beta = 1$. When only looking at the number of the 5 most relevant scenarios predicted by the reference that are in the top 10 predicted by the indicators, both approaches detect all 5 of them in each scenario.

5.4. Cross-validation of the data-driven approaches

While the two approaches that incorporate the demand side showed the best results for all tasks so far, they also both rely on training data from the power flow simulations that were used as reference. These results will, however, not be available in the typical use case where the indicators I assessed would be used for a pre-analysis. One way to solve this problem is to run power flow simulations for a few randomly selected scenarios and use the results to train the indicators. To assess how well the indicators perform when trained on such a subset of scenarios, I conducted a cross-validation on 100 subsets each with 5 randomly sampled weather years from the 31 weather years of the ERAA 2022 dataset. More precisely, I trained both indicators on each of these subsets including all technical scenarios and then predicted the hours with electricity shortages for the whole dataset.

Figure 6 shows box-and-whisker plots of the sensitivity and precision as well as the F_1 - and F_2 -score that the indicators trained on the subsets achieve. Each box shows the distribution of the values over each of the 100×12 combinations of a subset and a technical scenario. As before, each optimisation was done twice: With the goal to maximise either the F_1 -score or the F_2 -score.

Overall, the median values achieved for each of the 4 measures, 2 indicators and 2 optimisation goals do not change substantially compared to the case where the whole dataset was used for training. The biggest change can be seen for the precision of my indicator when optimised for the F_1 -score. It decreases by 0.017

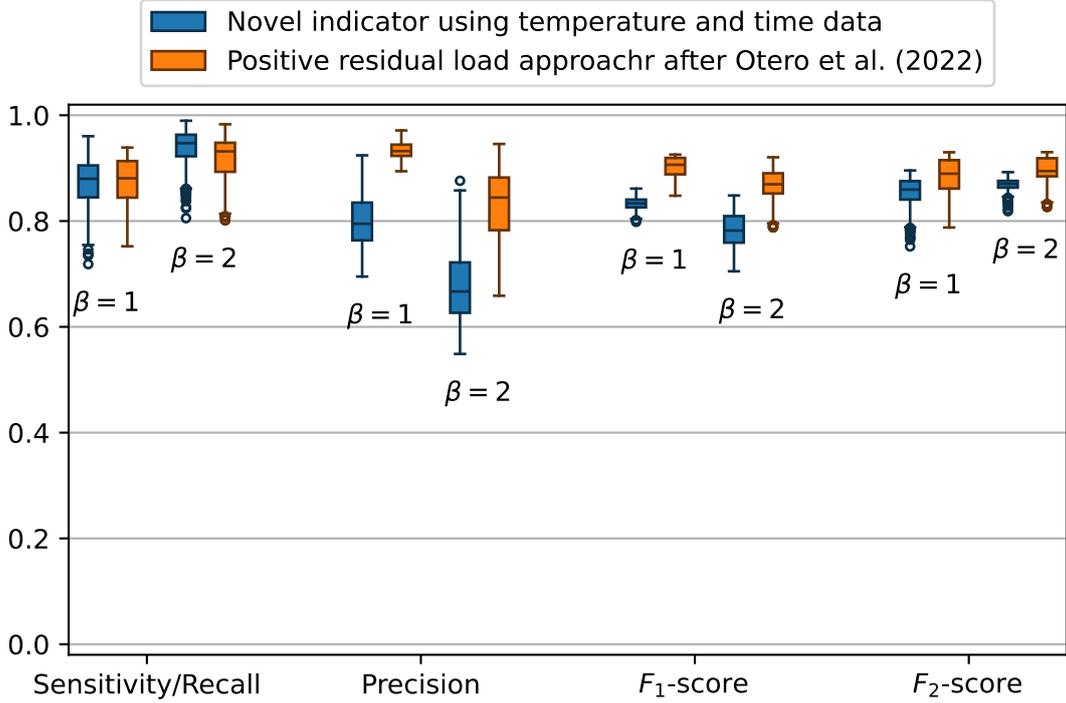


Figure 6: Box-whisker plots of the distributions of sensitivity, precision, F_1 - and F_2 -score of the 2 trained indicators in the cross-validation. The indicators are optimised for the F_β -score with two choices of beta: $\beta = 1$ and $\beta = 2$. The boxes show the distribution of the achieved values over 100×12 configurations of 100 random choices of 5 weather years for training years and the 12 scenarios in the ERA5 2022 dataset.

from 0.812 to 0.795. A greater deviation can be seen in the worst case performance. The changes generally stay below losses of 0.1 though, with only two exceptions. The approach after Otero et al. (2022) loses 0.159 of its minimal precision when optimised for the F_2 -score, dropping to 0.659, while the minimal precision of my indicator when optimised for the F_1 -score is decreased by 0.107 to 0.695. Looking at the F_β -score that the algorithms were optimised for the corresponding values did not fall below 0.798, which is still a good value compared with F_1 -scores below 0.5 attained in a similar analysis for daily time series by Biewald et al. (2025).

Analogously, Figure 7 depicts the distribution of the MAE and MASE for the results of the cross-validation. As the precision is most affected by the reduction

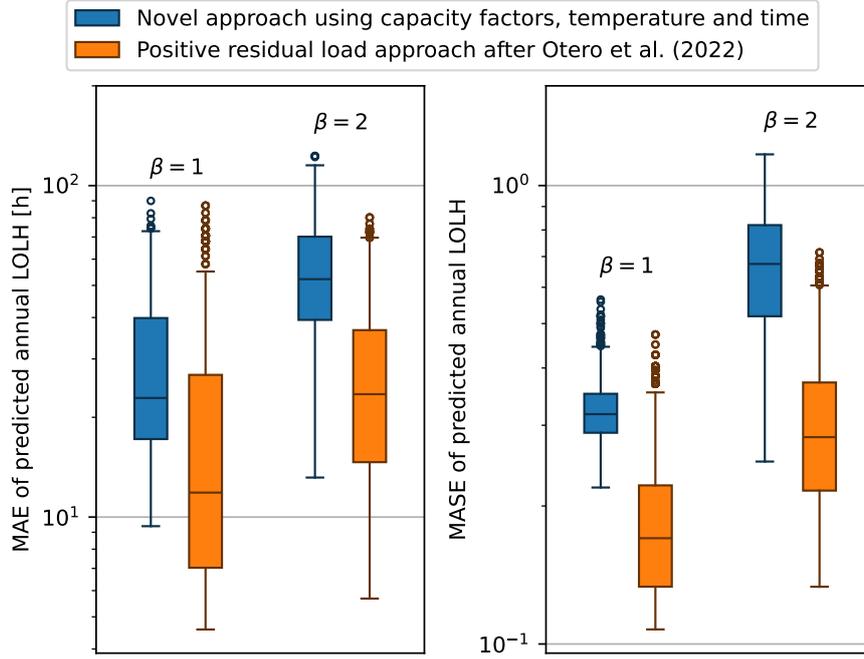


Figure 7: Boxplots of the MAE of the predicted annual LOLH of the two trained indicators in the cross-validation. The boxes show the distribution of the achieved values over 100×12 configurations of 100 random choices of 5 weather years for training and the 12 scenarios in the ERAA 2022 dataset. Each indicator was optimised for the F_β -score with two choices of beta: $\beta = 1$ and $\beta = 2$.

of the training data, a tendency to overestimate the LOLH can be expected. In fact, the impact on the MAE is greater than on the binary prediction measures. It increases by about 26.95% from 18.02 h to 22.87 h for my indicator when optimised for the F_1 -score and by about 25.84% from 18.66 h to 23.48 h for the residual load approach when optimised for the F_2 -score. The resulting absolute change in the median MASE stays below 0.032, though, for all four configurations. Again, the outliers of the distribution increased more, however, resulting in an increase of the maximal MAE by up to 42.74 h or 53.34% for my indicator when optimised for the F_2 -score.

The results for the rank correlation change only marginally with deviations of the median and minimum below 0.01 and 0.04 respectively (see Figure 8). Interestingly, the median Sørensen-Dice coefficient actually improved slightly from 0.95

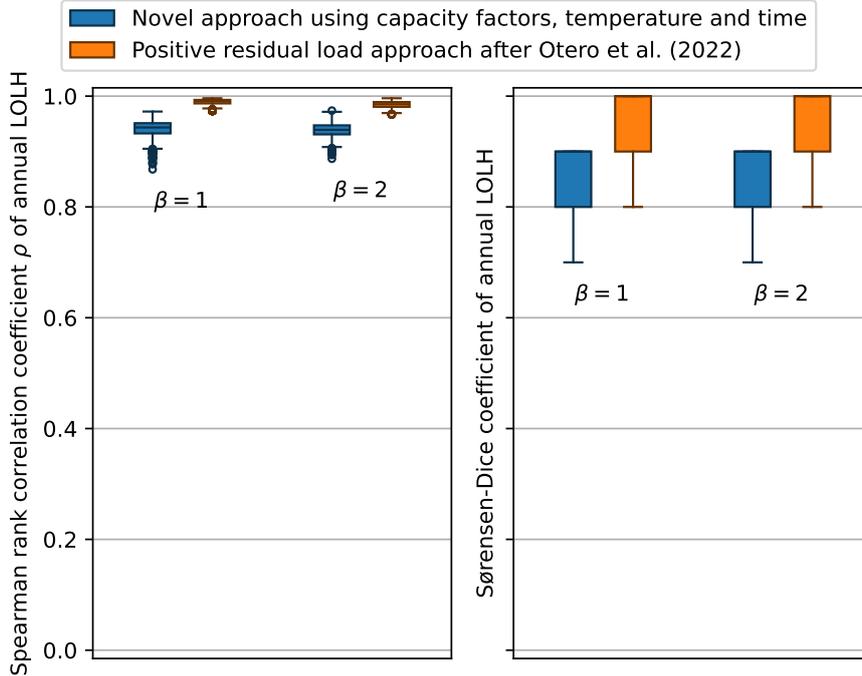


Figure 8: Box plots of the Spearman rank correlation coefficient (left panel) and the Sørensen-Dice coefficient (right panel) of the two trained indicators on the ERAA 2022 dataset in the cross-validation. The boxes show the distribution of the achieved values over 100×12 configurations of 100 random choices of 5 weather years for training and the 12 scenarios in the ERAA 2022 dataset. Each indicator was optimised for the F_β -score with two choices of beta: $\beta = 1$ and $\beta = 2$.

to 1 for the residual load approach after Otero et al. (2022), indicating that for more than half of the considered scenarios all of the ten weather years with the most annual LOLH in the reference were also ranked in the top ten positions by the indicator. The Sørensen-Dice coefficient for my indicator stayed the same in the median but got worse by 0.1 if trained for the F_2 -score. In fact my indicator only always identified all of the 5 weather years with the most LOLH in the reference as being within the most critical ten scenarios if trained for the F_1 -score.

Overall, the cross-validation shows that a training set of five random weather years suffices to train the two indicators and retain most of the predictive performance. More than that, the cross validation gives an indication about the appropriate choice for the parametrization of the F_β -score. While the analysis of

the indicators trained on the complete dataset suggest that the choice of β depends on the task at hand, the cross-validation show preferable results for $\beta = 1$ for all tasks.

5.5. Transferability to a different dataset

As another test of the generalisability of the approaches, I repeated the analysis for the ERAA 2024 data as a second dataset. As described in section 3.1.4, the data looks quite different from the ERAA 2022 data. Especially the redefinition of the demand, that incorporates out-of-market capacities as negative loads, poses a challenge for my indicator as it is structurally blind to these changes. While the approach after Otero et al. (2022) is aware of these changes, it is, like my indicator, ignorant toward changes in storage and transmission capacities as well as the availability of other electricity sources than on- and offshore wind and solar PV. Thus, the indicators have to be retrained on the new dataset to achieve reliable results. An analysis of the performance of the indicators trained on the ERAA 2022 data at predicting electricity shortages in the ERAA 2024 data is given in Section 5.6.

5.5.1. The exact prediction of hours with electricity shortages on the ERAA 2024 data

Analogously to Figure 2, Figure 9 shows the sensitivity, precision, F_1 - and F_2 -score of all five indicators achieved on the ERAA 2024 dataset. First looking at the three VRE drought definitions, an increased sensitivity can be seen compared to the results on the ERAA 2022 data. This high sensitivity is however a result of a vast overestimation of the annual LOLH, which leads to very low precision values below 0.025. The resulting poor results for the F_β -scores show the missing adaptability of the VRE drought definitions to different circumstances.

Unsurprisingly, the two indicators trained on the data show better results, although not on the level of the results for the ERAA 2022 data. My indicator has a slightly better sensitivity but significantly worse precision than the residual load approach after Otero et al. (2022), again due to an overestimation of the total amount of electricity shortages. The resulting median F_1 -scores of 0.597 and 0.507

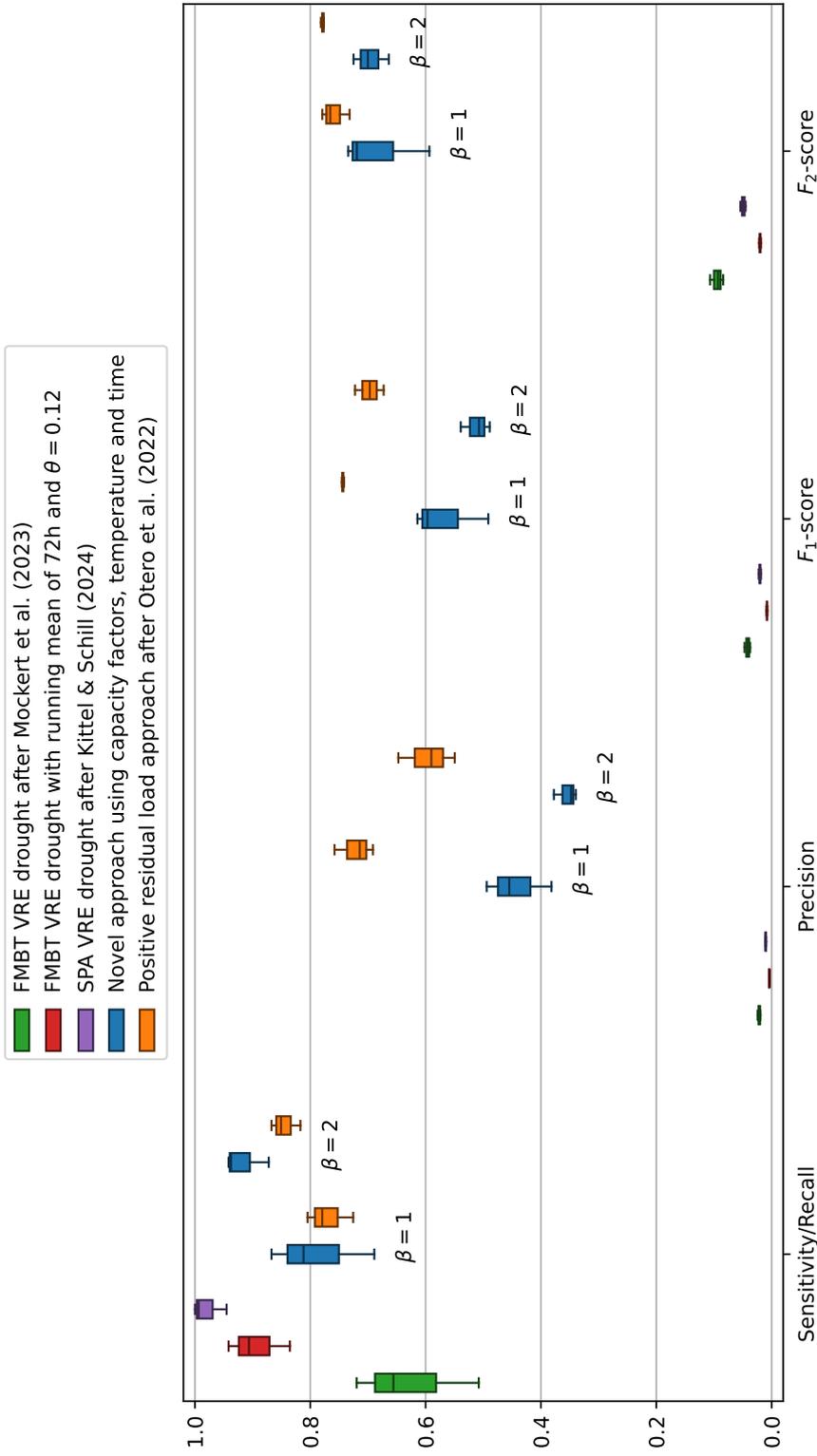


Figure 9: Box-whisker plots of the distributions of sensitivity, precision, F_1 - and F_2 -score of the five considered indicators on the ERAA 2024 dataset. The two trained indicators are optimised for the F_β -score with two choices of beta: $\beta = 1$ and $\beta = 2$. The boxes show the distribution of the achieved values over the 3 climate models in the ERAA 2024 dataset.

for an optimisation of the F_1 - respectively F_2 -score are still good compared to the results attained for daily time series by Biewald et al. (2025) in a similar analysis but 0.147 and 0.189 points worse than the results of the residual load approach. The values for the F_2 -score are more similar but show an interesting feature that seems to be contradictory at first sight. The F_2 -score of my indicator is better when the SVMs are optimized for the F_1 -score and not the F_2 -score. This is not a problem with the optimisation of the individual SVMs, but an effect of the lack of sufficient samples with an electricity shortage. 38 of the 48 subsets defined by the hour of the day and the classification in regular workdays and other days do not exhibit at least 5 samples with an electricity shortage. Therefore, no SVM can properly be trained on them and the SVM of another subset is used as described in Section 4. So, most of the samples are not classified by an SVM that is optimised on the appropriate subset. Thus, the F_2 -score can not be expected to be optimised. As the optimisation for the residual load approach is done on the whole dataset, it does not need as many positive samples to be trained properly.

Again, I conducted a cross-validation for which the indicators are trained on 100 subsets of the whole dataset, each consisting of five randomly sampled weather years, each with three scenarios from the three considered climate models. The results are depicted in Figure 10. The problem of too few positive training samples is exacerbated in the cross-validation. The sensitivity of both indicators only decreases by less than 0.06 in the median. However, the already low median precision of my indicator almost halves to only 0.199 in the case where it is optimised for the F_2 -score. The median F_1 -scores are accordingly low, but at 0.404, when optimized for the F_1 -score, still approximately at the level achieved with daily sampled residual load time series by Biewald et al. (2025). The changes in the minimal values were again bigger leading to worst-case F_1 -scores well below 0.2 for my indicator. The residual load approach after Otero et al. (2022) performs much better and actually gains 0.044 in its median precision, again showcasing the lesser dependence on the amount of positive training samples. The median F_β -scores stayed almost constant and even the minimal values do not drop below 0.691 if trained for the considered score. This can still be regarded to be a good performance compared to the results of Biewald et al. (2025), who used the same approach with daily time series.

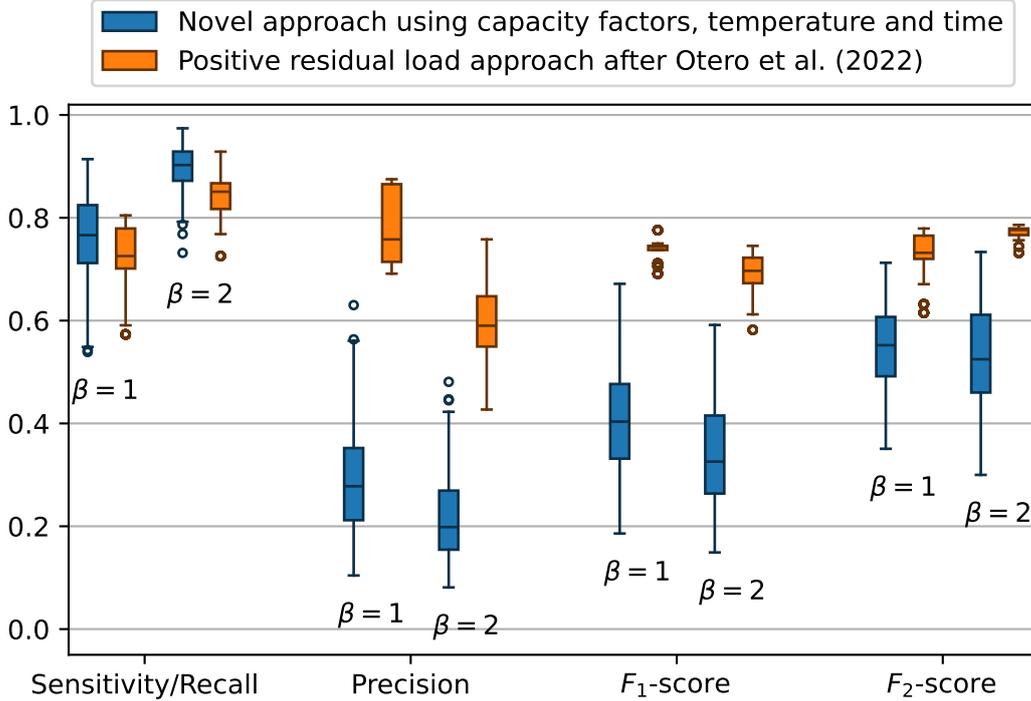


Figure 10: Box-whisker plots of the distributions of sensitivity, precision, F_1 - and F_2 -score of the 2 trained indicators in the cross-validation. The indicators are optimised for the F_β -score with two choices of beta: $\beta = 1$ and $\beta = 2$. The boxes show the distribution of the achieved values over 100×3 configurations of 100 random choices of 5 weather years for training and the 3 climate models in the ERAA 2024 dataset.

5.5.2. The prediction of annual LOLH on the ERAA 2024 dataset

Figure 11 shows the distribution of the MAE and MASE of the predicted annual LOLH for the ERAA 2024 data. As was indicated by their very low precision values, the three VRE drought definitions vastly overestimate the annual LOLH. While the power flow simulations predict 9 h of electricity shortages per year in the median, the VRE droughts predict between 355 h and 2685.5 h.

As before, the optimisation of the F_1 -score achieves better results at this task, so, I will focus my analysis on the corresponding values. When trained on all available data, my indicator performs much better than the VRE droughts with a median and maximal MAE of about 9.7 h and 12.5 h respectively. However, the

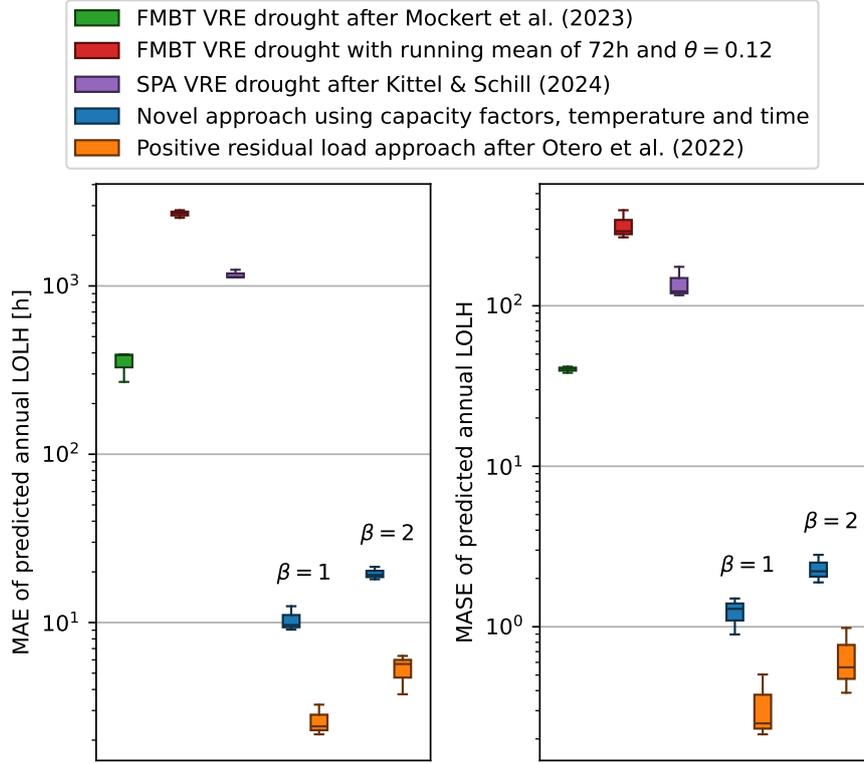


Figure 11: Box plots of the MAE and MASE of the predicted annual LOLH of the five considered indicators compared to reference results from power flow simulations for the ERAA 2024 data. The two trained indicators are optimised for the F_β -score with two choices of beta: $\beta = 1$ and $\beta = 2$. The boxes show the distribution over the three climate models of the ERAA 2024 dataset.

median MASE of 1.29 indicates that the constant prediction of the mean annual LOLH would achieve a better MAE.

The cross-validation, depicted in Figure 12, shows that my indicator predicts significantly higher levels of annual LOLH if only 5 years are used for training. This can be explained by the lack of sufficient examples of electricity shortages. The SVM are only trained for the hours of the day with sufficient amounts of electricity shortages, which are hours with above average electricity demand. Therefore, the available SVM will overestimate the electricity demand for the other hours of the day that they are assigned to. As a result, too many hours with electricity shortages will be predicted for these hours of the day. This suggests, again, that

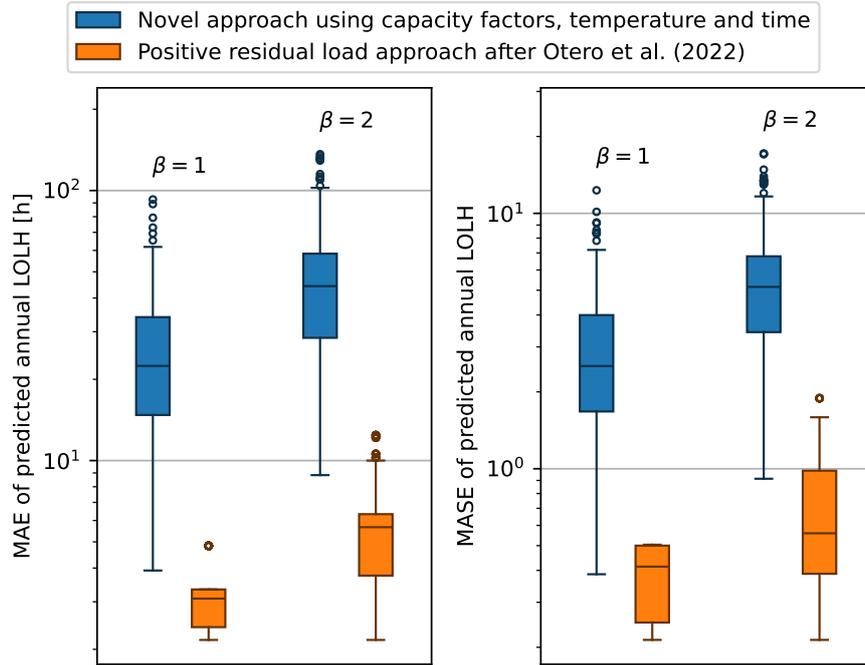


Figure 12: Box plots of the MAE (left panel) and MAE (right panel) of the annual LOLH for my indicator and the residual load approach on the ERAA 2024 dataset. The indicators were trained on 100 randomly chosen subsets of 5 weather years from the ERAA 2024 dataset. The optimisation was done for the F_β -score with two choices of beta: $\beta = 1$ and $\beta = 2$. The boxes show the distribution of the values achieved for the 3×100 combinations of climate models and training sets.

the problems of my indicator on the dataset mainly stem from the lack of positive samples to train a sufficient amount of the 48 SVMs.

The residual load approach after Otero et al. (2022) achieves a median and maximal MAE of about 2.4 h and 3.3 h respectively. The resulting maximal MASE of 0.50, that is not increased in the cross-validation, shows that the approach is still able to predict a good part of the year-to-year variability of the annual LOLH.

5.5.3. The prediction of the most critical weather years on the ERAA 2024 data

As before, I also assessed the ability of the indicators to predict the relative criticality of the weather years for resource adequacy concerns. Figure 13 shows the

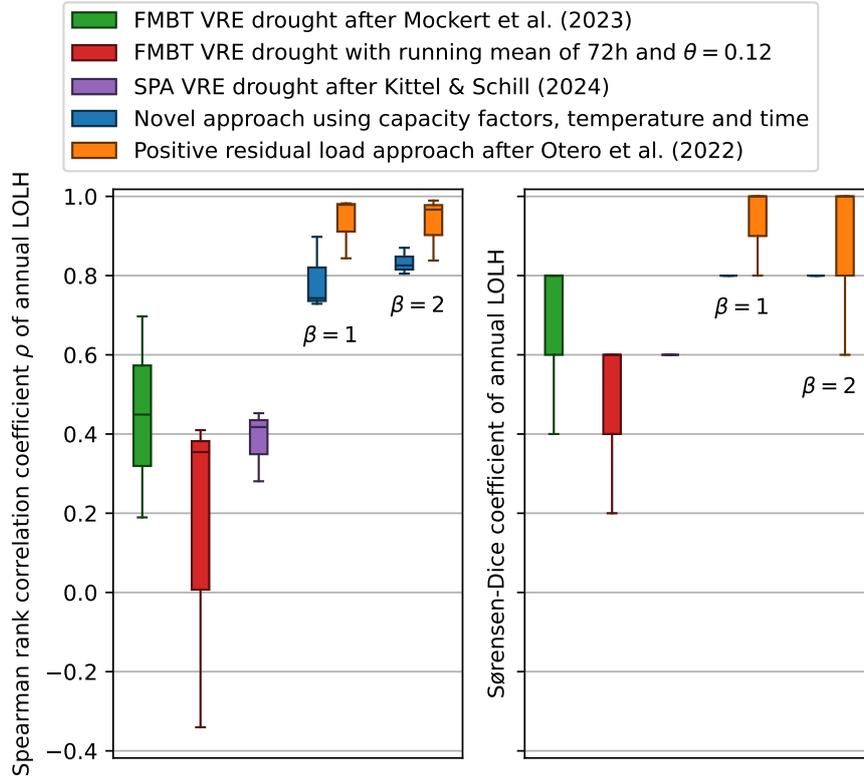


Figure 13: Box plots of the Spearman rank correlation coefficient (left panel) and the Sørensen-Dice coefficient (right panel) of the five considered indicators on the ERAA 2024 dataset. The two trained indicators are optimised for the F_β -score with two choices of beta: $\beta = 1$ and $\beta = 2$. The boxes show the distribution of the achieved values for the three climate models in the ERAA 2024 dataset.

distribution of the Spearman rank correlation ρ and the Sørensen-Dice coefficient for the ERAA 2024 data, analogously to Figure 5. Because only 12 weather years are available per climate model, here, the Sørensen-Dice coefficient of the 5 years with the highest annual LOLH predicted by the indicators and the reference is shown.

The VRE drought definitions, again, show rather poor results for the Spearman rank correlation. Notably, the long FMBT approach, that performed best at this task on the ERAA 2022 data, exhibits the lowest values here, with a minimum of -0.340 . The other two VRE droughts on the other hand have higher values than before for both the Spearman rank correlation and the Sørensen-Dice coefficient.

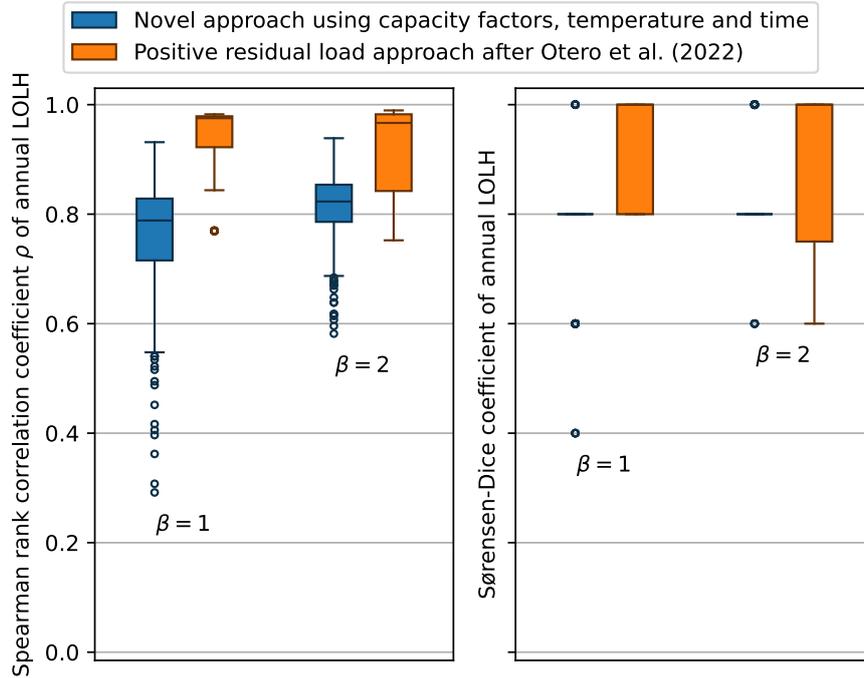


Figure 14: Box plots of the Spearman rank correlation coefficient (left panel) and the Sørensen-Dice coefficient (right panel) for my indicator and the residual load approach on the ERAA 2024 dataset in the cross-validation. The indicators are optimised for the F_β -score with two choices of beta: $\beta = 1$ and $\beta = 2$. The boxes show the distribution of the achieved values over 100×3 configurations of 100 random choices of 5 weather years for training and the 3 climate models in the ERAA 2024 dataset.

They actually detect all 3 weather years with the highest annual LOLH as being in the top 6 for all climate models.

The Spearman rank correlation of my indicator is clearly decreased compared to the ERAA 2022 data. With values between 0.728 and 0.898 its performance is however still acceptable which is confirmed by the Sørensen-Dice coefficient of 0.8 and its ability to predict the set of the 3 most critical weather years for all 6 configurations. However, the cross-validation shows that the performance can drop significantly if the training data is reduced to 5 years. The Spearman rank correlation in the worst case is only 0.292 and the Sørensen-Dice coefficient drops to 0.4 indicating that for some configurations only 2 of the 5 most critical weather years were ranked in the top 5. Consequently, the indicator is only able to predict

all 3 most critical weather years to be in the top 6 for 471 of the 600 configurations.

The residual load approach after Otero et al. (2022) again performs best, and while its worst case performance is lower than before its median Spearman rank correlation is still above 0.966 in the cross validation. Furthermore, it correctly classifies all 3 most critical weather years as being in the top 6 in all configurations of the cross-validation.

5.6. Transferability to another dataset without renewed training

As a final test of the transferability of the indicators, I trained my indicator and the residual load approach after Otero et al. (2022) on the complete ERAA 2022 dataset and applied them on the ERAA 2024 data without a renewed training.

5.6.1. The exact prediction of hours with electricity shortages on the ERAA 2024 data with pretrained indicators

Figure 15 shows box plots of the sensitivity, precision, F_1 - and F_2 -score of the two indicators. While my indicator achieves nearly perfect sensitivity, its precision is very low, with values between 0.060 and 0.089. Consequently, the resulting F_β -scores are unsatisfactory but still clearly preferable to the results of the three VRE drought definitions.

The residual load approach exhibits its lowest sensitivity values at 0.409, but achieves good precision values above 0.821. As a result, the F_β -score values are still relatively high. Interestingly, the indicator performs significantly better in both scores if it was optimized for the F_2 -score on the ERAA 2022 data. In this case it has F_1 -scores between 0.690 and 0.767, well above the ones achieved by Biewald et al. (2025).

The poor performance of my indicator compared to the residual load approach is partly due to the redefinition of the electricity demand in the ERAA 2024 that leads to much lower demand values for otherwise similar circumstances. Consequently, the indicator overestimates the demand and therefore the severity of a situation as it learned the dependence of the demand on temperature and time in the old

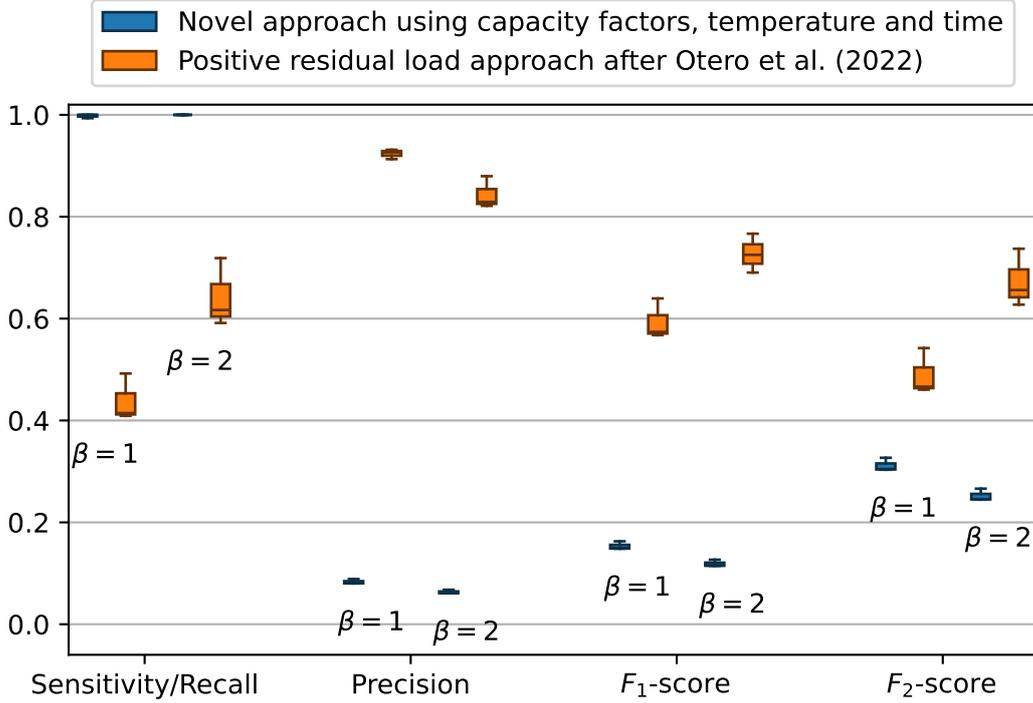


Figure 15: Box-and-whisker plots of the distributions of sensitivity, precision, F_1 - and F_2 -score of my indicator and the residual load approach on the ERAA 2024 data. Both were pretrained on the ERAA 2022 dataset with an optimisation of the F_β -score for two choices of beta: $\beta = 1$ and $\beta = 2$. The boxes show the distribution of the achieved values over the 3 climate models in the ERAA 2024 dataset.

dataset. The residual load approach on the other hand has access to the demand data directly and therefore automatically adjusts to the redefinition.

5.6.2. The prediction of annual LOLH on the ERAA 2024 data with pretrained indicators

Figure 16 shows box plots of the MAE and MASE of the prediction of the annual LOLH as achieved by the two pretrained indicators on the ERAA 2024 data. As could be expected from the low precision values, my indicator performs poorly with MAE and MASE values above 121.7h and 12.91, respectively. The residual load approach performs much better, especially, when it was trained to optimize the

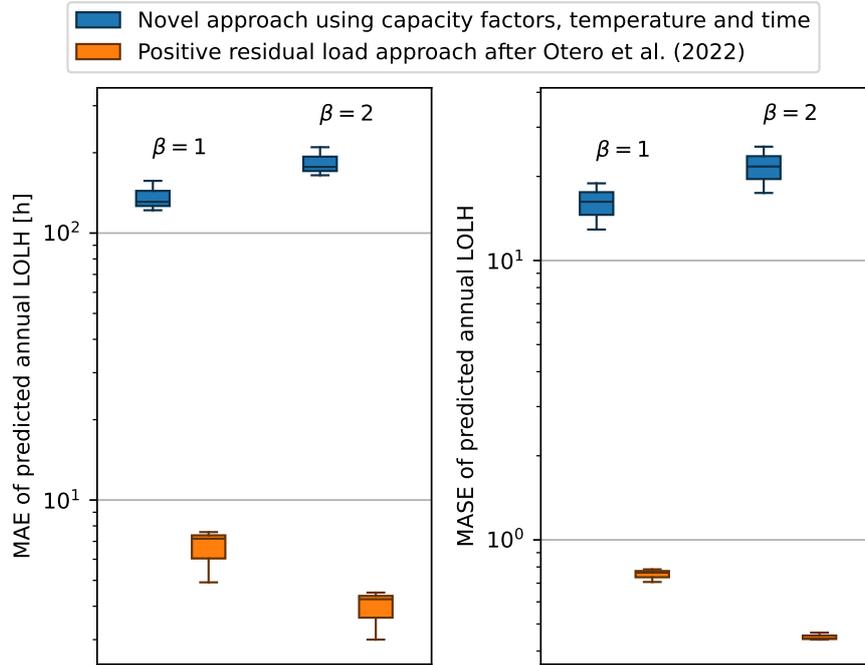


Figure 16: Box plots of the MAE (left panel) and MASE (right panel) of the predicted annual LOLH for my indicator and the residual load approach on the ERAA 2024 dataset. Both indicators were pretrained on the ERAA 2022 dataset with an optimisation of the F_β -score with two choices of beta: $\beta = 1$ and $\beta = 2$. The plots show the distribution over the 3 considered climate models.

F_2 -score. In this case it has a MAE and MASE below 4.5 h and 0.47, respectively. So, the predictions are still twice as accurate as always predicting the mean, which would not even be possible without training data.

5.6.3. The prediction of the most critical weather years on the ERAA 2024 data with pretrained indicators

Finally, Figure 17 depicts box plots of the Spearman rank correlation and the Sørensen-Dice coefficient, analogous to Figure 13. My indicator achieves a Spearman rank correlation between 0.596 and 0.816 indicating that it is able to rank the weather years reasonably well. With a Sørensen-Dice coefficient between 0.6 and 0.8 it is however only able to predict 3 to 4 elements of the set of the 5 weather

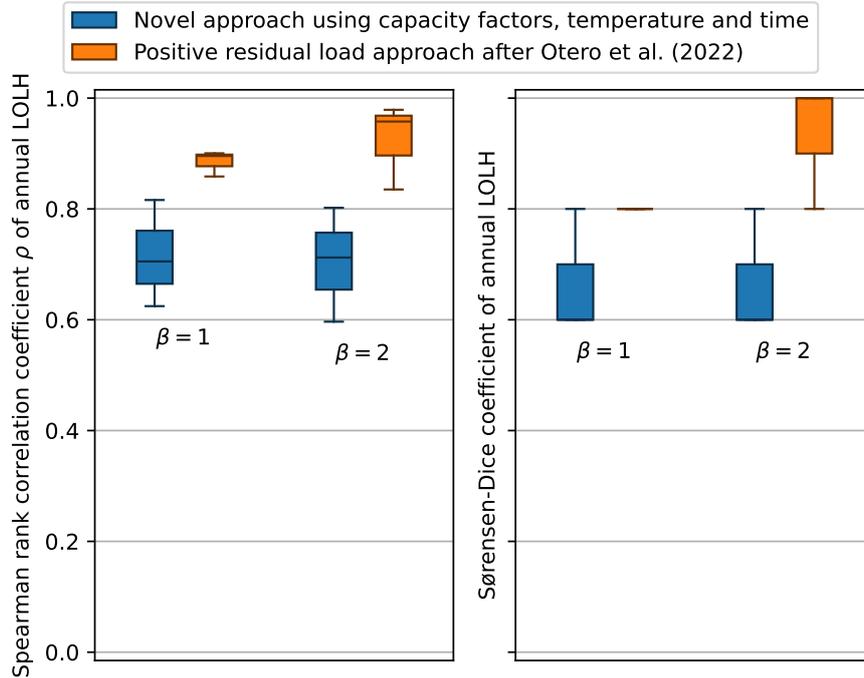


Figure 17: Box plots of the Spearman rank correlation coefficient (left panel) and the Sørensen-Dice coefficient (right panel) for my indicator and the residual load approach on the ERAA 2024 dataset. Both indicators were pretrained on the ERAA 2022 dataset with an optimisation of the F_β -score with two choices of beta: $\beta = 1$ and $\beta = 2$. The plots show the distribution over the 3 considered climate models.

years with the highest annual LOLH. Furthermore, it successfully identifies the 3 most critical weather years as being within the 6 most critical ones in only 4 of the 6 considered configurations.

The residual load approach, again, performs better, achieving a rank correlation between 0.835 and 0.979, with the minimum value taken if the indicator was trained to optimize the F_2 -score. However, this optimization achieves a higher Sørensen-Dice coefficient in the mean and enables the indicator to identify all 3 of the most critical weather years as being within the 6 most critical ones unlike the optimisation for the F_1 -score.

6. Discussion

In the previous section I assessed the ability of the five considered indicators to predict electricity shortages on three different levels of accuracy: The prediction of exact hours at which electricity shortages occur, the prediction of annual LOLH and the identification of the climate years with the highest annual LOLH. To evaluate the sensitivity to the choice of training data, a cross-validation was conducted for the two data-informed indicators. Furthermore, the whole process was repeated for a different dataset that has different characteristics. Finally, the two data-informed indicators were trained on the first dataset and applied to the second in order to evaluate the ability to adapt to new circumstances without a renewed training. In the following, I will discuss the strengths and limitation of the five methods as predictors of electricity shortages based on the previous observations.

6.1. Discussion of the VRE drought indicators as predictors of electricity shortages

The three VRE drought indicators, that only consider the supply side of the electricity balance, show unsatisfactory results on all three levels of accuracy. As pointed out in section 5.1.1, this can be explained by their structural ignorance of the electricity demand. They can not distinguish between a dark doldrum on a warm Sunday night in June, where the electricity demand can be expected to be low, and a dark doldrum on a cold Tuesday evening in January, where electricity demand will typically be much higher. Therefore, they either have to predict an electricity shortage for comparatively high capacity factors in order to detect situations like the aforementioned Tuesday evening in January at the price of false alarms in warm summer nights or they do not detect either. This is reflected in the results achieved by the indicators. The FMBT approach after Mockert et al. (2023) has a low sensitivity compared to the other two VRE drought definitions, so it misses a lot of electricity shortages, but, on the other hand, it does not produce as many false alarms. Overall, all three definitions achieve similar, low, F_β -scores, showing that their ability to prioritise sensitivity or precision over the other is not able to overcome their structural shortcomings. This is in accordance with the

results of Biewald et al. (2025), who varied the threshold parameter of a similar CBT VRE drought definition for daily sampled time series and achieved F_1 -scores below 0.3 for all parameter values.

6.2. Discussion of the residual load approach as an indicator of electricity shortages

The residual load approach after Otero et al. (2022) achieves the best results of the five indicators for all considered levels of accuracy and datasets. The reason for this success is likely to be found in the direct use of the residual load. To better understand how exactly that helps to predict electricity shortages and under which circumstances the indicator might be less effective, it is insightful to look at the energy flow balance:

$$E_{VRE} + E_{oRES} + E_{conv} + E_{store} + E_{imp} = E_{load}, \quad (23)$$

where E_{VRE} , E_{oRES} , E_{conv} , E_{store} , E_{imp} and E_{load} are the electrical energy provided, or taken up by VRE energy sources, other renewable energy sources, conventional electricity generators, storage options, the net import from neighbouring countries and the electrical load, respectively.⁵ Now, by definition, electricity shortages occur exactly when Equation (23) does not hold. An approach that uses the residual load can then simplify Equation (23) significantly by reshuffling the terms to get:

$$\text{residual load} = E_{load} - E_{VRE} = E_{oRES} + E_{conv} + E_{store} + E_{imp}. \quad (24)$$

As the residual load is assumed to be known, this removes E_{load} and E_{VRE} from the equation. And these two variables are especially hard to predict: For one, they vary a lot over time. Additionally, the control of E_{load} in the form of demand response is very limited, and E_{RES} can be reduced at will but not increased. An oversupply of energy can be handled easily most of the time, e.g., by shutting down generators, and can be neglected as cause of energy shortages here because the power flow simulations used as reference assume that all generators can be

⁵Note, that only on- and offshore wind and solar PV are considered to be VRE sources here, as in the rest of this thesis.

turned off instantaneously. Therefore, it is valid to focus on the situation where the right hand side of Equation (24) falls short of the residual load. The residual load approach after Otero et al. (2022) now simply assumes that this is always the case if a constant level of the residual load is crossed. So, why does this assumption work so well? The conventional power plants are modelled to be always and instantaneously available with their full nameplate capacity. Thus, they always max out at a constant level, which only leaves E_{oRES} , E_{store} and E_{imp} as sources of variation. Apparently, these three variables either usually balance out quite well or the variations of their maximal supply in critical situations are small compared to the critical residual load in the considered datasets. Therefore, the residual load approach can be expected to show a good predictive performance if it has sufficient training data to calibrate its threshold and the variation of the maximal output of E_{oRES} , E_{store} and E_{imp} in critical situations does not increase compared to the critical residual load. In the upcoming decades the share of other renewable energy sources, such as hydro power, in the energy mix is not expected to grow in Germany (Spänhoff, 2014). However, storage solutions and transmission capacities are planned to be increased significantly and can play a crucial role in reducing the risk of electricity shortages (ENTSO-E, 2024b; Hagspiel et al., 2018; Luburić et al., 2018; Solar Power Europe, 2025). An analysis of the sensitivity to changes in the available transmission and storage capacities by means of power flow simulations could estimate the reliability of the indicator for the analysis of power systems further in the future and could be a worthwhile endeavour.

Apart from the better results, the residual load approach has the advantage of needing less data samples for training than my indicator, as it has only one parameter. This is confirmed by the comparatively small impact of the reduction of training data in the cross-validation. The optimisation of the F_β -score in the calibration of the threshold, generally showed better results for the F_1 -score, that gives equal weights to precision and sensitivity, compared to the F_2 -score. For tasks, where either precision or sensitivity is more important than the other, another choices of β might still be more appropriate though. For a higher weight on the sensitivity, so an increased β , this will move down the threshold and thereby increase the amount of detected electricity shortages. Analogously, smaller values of β will move up the threshold.

This also explains why the indicator trained for the F_2 -score on the ERAA 2022 dataset showed better results on the ERAA 2024 dataset than the one trained for the F_1 -score. The optimisation on the 2024 dataset yielded thresholds for $\beta = 1$ (100.4 GWh) and $\beta = 2$ (98.7 GWh) that were lower than both thresholds attained on the ERAA 2022 data (104.1 GWh for $\beta = 1$, 101.9 GWh for $\beta = 2$). Therefore, the lower threshold on the old dataset, corresponding to $\beta = 2$ as explained above, was closer to the optimal values of the threshold for the F_1 - and the F_2 -score on the ERAA 2024 dataset. However, this would be the other way around if I had used the indicator trained on the ERAA 2024 dataset on the ERAA 2022 data. Consequently, if no information is available that could help decide if the optimal threshold of a new dataset is higher or lower than the one the indicator is trained on, I still suggest to optimise the F_1 -score. However, in many cases an indication of where the threshold should move might be available. For example, if the capacity of conventional power plants is changed, the threshold can be expected to be changed accordingly. Again, an analysis of the impact of a change of available capacities that are not considered in the residual load by means of power flow simulations would be beneficial. It could provide a rule of thumb of how to move the threshold if the dataset is changed and no training data is available.

Concluding, the residual load approach optimised for the F_1 -score seems to be an excellent choice for the fast evaluation of resource adequacy concerns if power flow simulations are not an option. It attains the best results of all indicators considered in this thesis. Additionally, the attained F_1 scores are about twice as high as the ones reported by Biewald et al. (2025) with the same approach but daily time series. So, the use of hourly time series is to be preferred if they are available. Of course, the residual load approach has one caveat: It requires the availability of load estimations. If this is not the case, my indicator is a good alternative.

6.3. Discussion of the novel indicator

My indicator performs almost as good as the residual load approach on the ERAA 2022 dataset, which included sufficient samples of electricity shortages to train all 48 SVMs corresponding to the combinations of the 24 hours of the day and

2 types of days. As claimed earlier, this works so well without any direct data on the electricity demand because the temperature and time are good proxies of the demand in the considered datasets. This can be understood by looking at the relationship of the temperature and the electrical load.

Panel (a) of Figure 18 shows a scatter plot of all data points from the ERAA 2022 dataset with the mean temperature on the horizontal and the load on the vertical axis. Although, a broad range of loads occur for a given temperature, a clear negative trend can be seen for lower temperatures that plateaus at about 10°C . This shape, a linear decrease for low temperatures and a constant plateau for higher temperatures, is a typical model for the relationship of load and temperature, also called Temperature Response Function (TRF) (Hu et al., 2024). In general, this model would predict increasing loads again after some higher temperature due to a demand for space cooling. While this trend is detected for other countries such as Greece and Italy, no such effect was recorded in Germany in the Summer of 2022 (Hu et al., 2024). This is not too surprising considering the relatively low availability of cooling devices such as Air Conditioner (AC) units in the country (Pezzutto et al., 2016). It is however interesting to note that no such trend can be seen in the demand data from ENTSO-E for 2030 depicted on Figure 18. The TSOs apparently assume no significant change in the cooling demand until 2030, although they assume an increasing penetration rate of heat pumps that would provide consumers with an effective tool for cooling their buildings (ENTSO-E, 2022, p. 64; Luo et al., 2015).

In order to utilise the TRF encoded in the data, it helps to isolate the other sources that influence the electricity demand. One important such driver is given by the hour of the day. Panels (b)-(e) of Figure 18 show the data split by the hour of the day, again as scatter plots with the same axes as in Panel (a) for four exemplary hours: 6am, 12am, 6pm and 12pm. Additionally, the data points stemming from different types of days are distinguished by colour. The three considered types of days considered here are: regular workdays, Saturdays & national German holidays and Sundays. Looking at the Panel with data from 6am for each day of the week the relationship between temperature and demand can now be described with a TRF as above but with a lot less noise. Unsurprisingly, the demand is lowest on Sundays, a bit higher on Saturdays and highest on regular weekdays. At

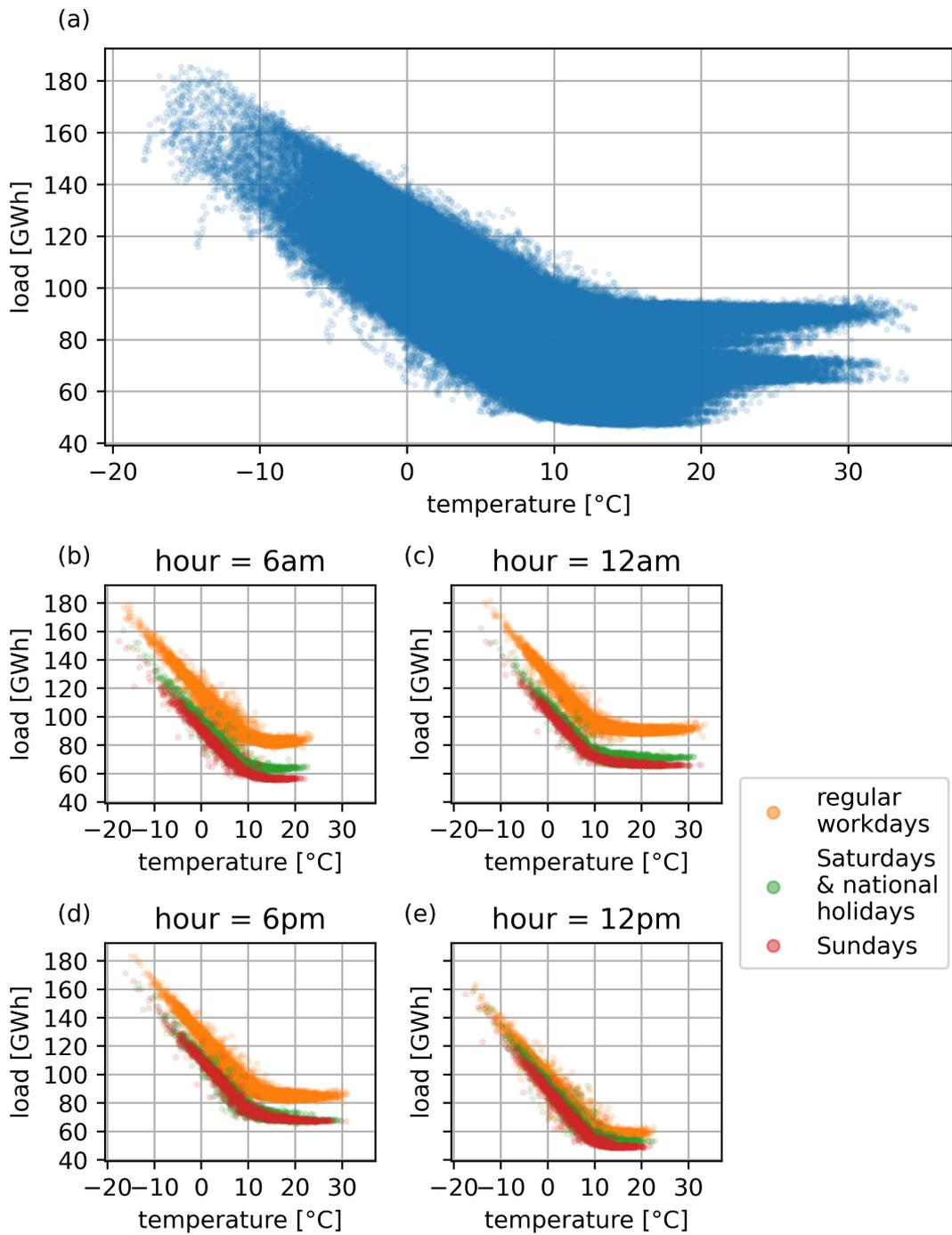


Figure 18: Scatter plot of ERAA 2022 data with the temperature on the horizontal axis and the electrical load on the vertical axis. Panel (a) shows all data points. Panels (b)-(e) show the data samples corresponding to one hour of the day each. Different day types are distinguished by colour.

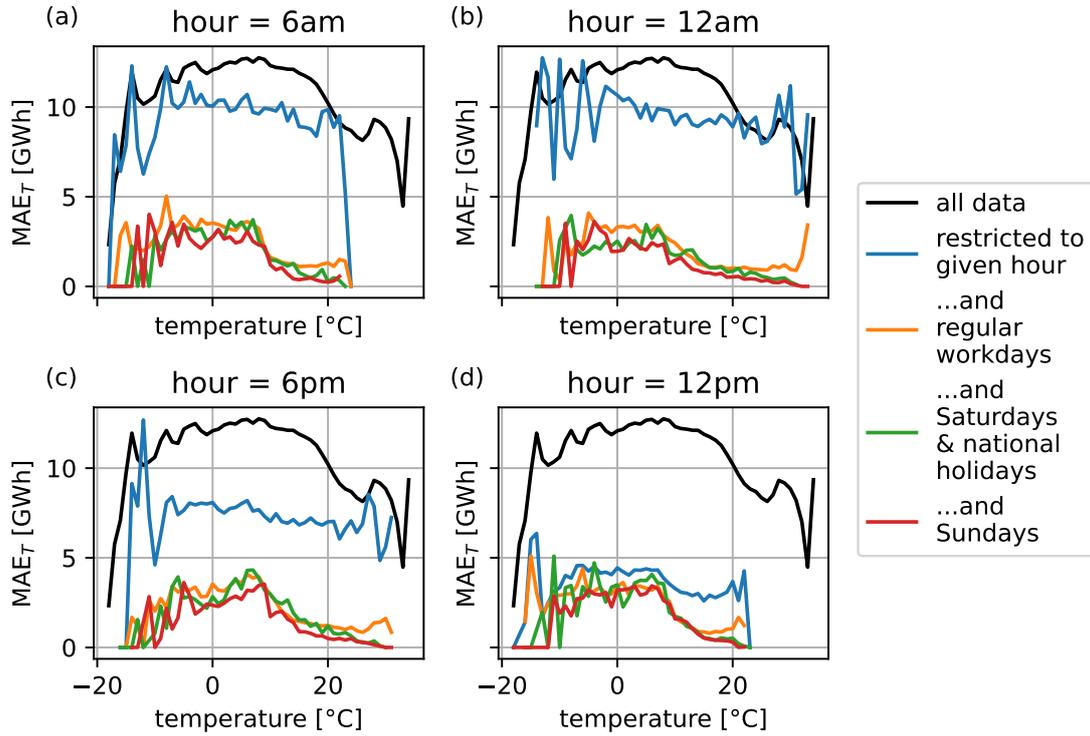


Figure 19: MAE_T as a function of the temperature T , where MAE_T is the MAE of the approximation of the residual load of a data samples by the mean residual load of all data points with the same rounded temperature. The black line in all panels shows the MAE_T for the whole ERAA 2022 dataset. Each panel then shows the standard deviation for the data at a given hour of the day in blue. The standard deviation for an additional restriction to regular workdays, Saturdays or national holidays and Sundays are depicted in orange, green and red, respectively.

12am the overall demand is higher and the constant part of the TRFs is longer as higher temperatures are more prevalent. At 6pm the average demand is still high at 95.2 GWh although it is lower in the constant part of the TRF for regular workdays. At 12pm the overall demand is low and weekdays, Saturdays and Sundays are much closer to each other.

The temperature can now be used to approximate the load by means of the observed distribution of loads for similar temperatures. For this analysis, I grouped the samples by rounding their temperatures to the next integer. This ensures the availability of sufficient samples for each group. The mean of the residual loads

for a given temperature can now be used as an approximation of the residual load for samples with that temperature. The reliability of this approximation can be assessed by the mean absolute error (MAE) for each group of samples with a given temperature T , which I denote by MAE_T . Figure 19 shows MAE_T of the approximation of the residual load as a function of the temperature in black. While it is low for very low temperatures, it increases steeply to about 12 GWh and fluctuates around that level for temperatures between -15°C and 15°C . For higher temperatures it declines to about 7 GWh. The overall MAE of an estimation of the residual load is then given by the average of MAE_T weighted by the amount of samples available for each temperature. To put this into a formula, let r^i and T^i be the residual load and the temperature of the i -th data sample, N and N_T the overall number of samples and the number of samples with temperature T and \bar{r}_T the mean of the residual load for samples with temperature T . Then, the overall MAE is given as

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |r^i - \bar{r}_{T^i}| = \frac{1}{N} \sum_T \sum_{i \in \{i | T^i = T\}} |r^i - \bar{r}_T| = \frac{1}{\sum_T N_T} \sum_T N_T \text{MAE}_T. \quad (25)$$

For the whole ERAA 2022 dataset, a value of approximately 11.9 GWh is attained. The four panels of Figure 19 show MAE_T for the subset of samples for four exemplary hours of the day in blue. It can be seen, that for all hours the estimation with the reduced set has a lower error. While the improvement is not too high at noon, the error is less than half as big as in the general case at midnight. This is in agreement with the better overlap of the TRFs for weekends and weekdays during the night. Averaging over all hours, the overall mean absolute error is about 7.3 GWh for the data split by the hour of the day, which is only approximately 61% of the error without this split. The effect of the additional separation into regular workdays, Saturdays & national holidays and Sundays is depicted by the orange, green and red graphs, respectively. Especially during the day, where the difference between regular workdays and weekends is more pronounced, the MAE of the further restricted data sets is significantly smaller, yet again. Using this split, the overall MAE of an estimation of the load given the temperature, hour of the day and type of day is only about 2.2 GWh for the ERAA 2022 data.

Unfortunately, the use of this split for my indicator would require the training of 72 SVMs: One for each combination of the 24 hours of the days and the 3 day types. The exacerbated problem with too few samples of electricity shortages for the training of the SVMs is the reason I decided to only differentiate between regular workdays and other days. The corresponding split of the data allows for an estimation of the electrical load with an overall MAE of about 2.5 GWh on the ERAA 2022 data. This is still only 34% of the MAE obtained if only the hour of the day and temperature are considered and 21% of the MAE that has to be expected if only the temperature is used. For very low temperatures, -10°C and below, where electricity shortages are particularly abundant, the MAE is even a bit lower at about 2.4 GWh for the separation into hours of the day and regular workdays and other days.

Having the relationship of electric load and temperature and time variables in mind, the effectiveness of my indicator on the ERAA 2022 dataset can be understood. In the dataset, electricity shortages only occur for temperatures below 8°C in Germany. Thus, only the linear part of the TRF is of importance. So, the electrical load can roughly be seen as a linear function of the temperature.⁶ If we now look at the space the SVMs are trained on, which is spanned by the combined capacity factor and the temperature, the data points with equal residual load will roughly be situated along a straight line or, mathematically speaking, a 1-dimensional hyperplane in \mathbb{R}^2 . Fittingly, the decision boundaries that linear SVMs learn are exactly hyperplanes. So, if my approach learns one of the hyperplanes of roughly equal residual load, it basically reduces to the residual load approach after Otero et al. (2022) with an additional error due to the uncertainty of predicting the residual load. Therefore, all sensitivities of the residual load approach discussed in section 6.2 apply also for my indicator. Of course, my indicator will not exactly learn such a hyperplane, but considering the success of the residual load approach it can be expected to learn a similar one if sufficient training data is available. Unsurprisingly, my indicator generally also provides better predictions if it is trained to optimise the F_1 -score instead of the F_2 -score, which therefore should be the standard. However, by choosing a different F_{β} -score for the optim-

⁶Here, linear function refers to a polynomial function of degree 1, not a linear map as defined in linear algebra.

isation, it is capable to adapt well to a situation where precision or sensitivity are prioritised without compromising the F_1 -score significantly.

Apart from the sensitivities to changes in the available capacities for other renewable sources, energy storage and transmission discussed for the residual load approach, my indicator depends on the relationship of the temperature and time variables with the electricity demand that it learns. So, if this relationship is changed significantly, it will need renewed training. This was the case for the ERAA 2024 dataset, for which the definition of the load was changed to include non-market batteries and solar PV. Additionally, like all indicators that rely on capacity factors instead of capacities, my predictor is blind to changes in the overall installed capacities. It could even happen, that my indicator predicts more electricity shortages when capacities are added if these capacities come from a technology with relatively low capacity factors, such as solar PV.

Probably the greatest weakness of my indicator is its dependence on sufficient samples of electricity shortages for each of the subsets that an SVM is trained on. As described in Section 5.5.1, these were not available for the ERAA 2024 dataset. In such cases, it might be advisable to use other methods. For example, the prediction of the annual LOLH had a MASE above 1 in the median, indicating, by the definition of the MASE, that the constant prediction of the mean annual LOLH would have a better MAE than my indicator. As this was even the case when my indicator was trained on the whole dataset, the prediction of the mean might be a better option in this situation. The reliability of my indicator if not all of its SVMs are trained can be estimated by assessing its predictive performance on the available training data.

A future improvement of my indicator could be achieved by reducing the number of its parameters to decrease its need for training data. One option would be to not train a different SVM for each subset but transform the input data according to a pre-defined function that takes the hour of the day and type of day into regard and then feed the transformed data to a single SVM. For instance, the temperature could be moved up and down to account for a typical hourly load curve of the type of day. Such a function could either be parametrized itself, which would make the optimisation process more challenging, or it could be fixed. In the case of a parameter-free function the overall number of parameters would have been

reduced from 96 (2 for each SVM) to 2. This would allow the one remaining SVM to be trained on the whole training data, which would mean that the approach could be trained with a comparative amount of data as the residual load approach.

Summarised, my approach is a valuable alternative to the residual load approach after Otero et al. (2022) for analyses involving electricity shortages if no load data is available. Due to its ability to approximate the electricity demand by means of temperature and time data, it achieves results that are almost as reliable as the ones of the residual load approach. However, care should be taken when training data is scarce as the indicators predictive performance is severely compromised if not all SVMs are optimised.

7. Exemplary application

Before concluding my thesis in Section 8, I will demonstrate an exemplary application of my indicator that highlights its utility for tasks that can not be assessed by the residual load approach.

7.1. The switch from reanalysis based data to climate projections in the ERAA 2024 and its impact on resource adequacy concerns

The consideration of potential effects of climate change on resource adequacy concerns is one of the methodological requirements of the ERAA (ACER, 2020, Article 4.1. (f)). Up until 2023, this was attempted by detrending the temperature of historical climate data and adjusting it to match the temperature expected in the target year (ENTSO-E, 2025a, pp. 56–57). However, this preliminary method is considered to be insufficient as it neglects other potential effects for example on the wind (ENTSO-E, 2025c, p. 64). Therefore, the ERAA 2024 used data from the 3 climate models included in the PECD 4.1. To ensure that the climatological conditions are representative of the target years, only the 12 surrounding years from 2025 to 2036 were used for the analysis (ENTSO-E, 2025c, p. 66). However, the impact of this change of input data on adequacy concerns can not easily be evaluated using the ERAA results alone, especially compared to other

uncertainties such as technical specifications of generators. The ERAA 2024 uses only the climate projections and not the historical data, and a comparison with earlier ERAAs is not immediately possible either because their assumptions about other parameters, e.g., the available infrastructure, are also different. M. Koivisto et al. (2023) studied the differences between reanalysis data and projections in an earlier version of what later became the PECD 4.1. They focused on the availability of solar PV and found the effect of climate change to be smaller than that of a change in the technical specifications of available wind generators. The general impact of climate change on adequacy concerns has also been examined (e.g. Ho et al., 2023), but the effect of the choice of climate years from either reanalysis or climate projections on resource adequacy assessments with a time horizon of less than 10 years is another question, that seems to be open for now. Therefore, an exploration of this question with the resource adequacy indicators evaluated in this thesis can be of interest for the energy systems modelling community. Since time series of the electricity load are only publicly available for 3×12 years of the projections, the residual load approach is not an option for this task. Thus, it seems to be the perfect task for an exemplary application of my indicator.

7.1.1. Analysis of the differences between the reanalysis and climate projection based datasets in the PECD 4.1

For this analysis, I used a total of 5760 years of climate data: reanalysis data for the 42 weather years from 1980 to 2021 and projections for the 50 weather years from 2015 to 2064 from all three climate models of the PECD 4.1 combined with all combinations of the 3 offshore and 10 onshore technical wind scenarios. As the focus of my analysis was the difference between the results from reanalysis and projections, the total installed capacities were of secondary interest and I decided to use the assumptions of the PEMMDB for the ERAA 2022. This allowed me to use the version of my indicator that was trained on the ERAA 2022 dataset for the prediction of electricity shortages. Based on my discussion in Section 6, it is reasonable to believe that the problems of my indicator with the ERAA 2024 dataset were mainly due to the changes in assumptions about the installed infrastructure and especially the redefinition of the electricity demand. Therefore,

I assume that my indicator has a predictive performance comparable to the one on the ERAA 2022 dataset in this analysis. However, it would be interesting to evaluate this by comparing with results of power flow simulations for the same configurations. Unfortunately, this was not possible to me because I did not have access to the corresponding demand data.

Figure 20 gives a first idea of the datasets. It shows the annual full load hours and LOLH for the 30 technical scenarios and climate data from the reanalysis and the MEHR model. The results for the other two climate models look similar in the aspects I want to point out in the figure. For both variables the transition from reanalysis data to the projections is not obvious at first glance. Looking at the full load hours, it is clear that the choice of the technical scenario for the onshore wind generators has a much higher impact than the choice of the offshore scenario or the decision to use climate projections or reanalysis data. Interestingly, the onshore scenarios 31, 35 and 39 show very similar results. For the predicted LOLH, the situation is not so clear as the overall data level is mostly dominated by a great variation between the years. Figure 21 shows the same situation but with a running mean of 10 years applied to the time series. It becomes clear that the choice of the onshore wind scenario also has the highest impact on the predicted LOLH. More than that, there seems to be a decreasing trend over the years in the reanalysis data. The running mean of the annual full load hours shows a difference of the general data level between the historical data and the projections. This is, however, not the case for all three models. In the following, I will assess this difference more rigorously with particular attention to the variation given by the three climate models.

In order to see the differences between the reanalysis dataset and the projection dataset and not between the scenarios, I calculated the mean values over the years of both variables for each of the scenarios and each dataset separately. Then I studied the percentage change of these means for the change from the reanalysis data to the projections, so the percentage of increase or decrease from the reanalysis data to the projections. To put this into a formula for the LOLH, let $LOLH_{model,on,off}^{year}$ be the annual LOLH of the year specified by the superscript for the climate model or reanalysis defined by *model* and the on- and offshore wind scenarios defined by *on* and *off*. Then, the percentage change of the mean annual

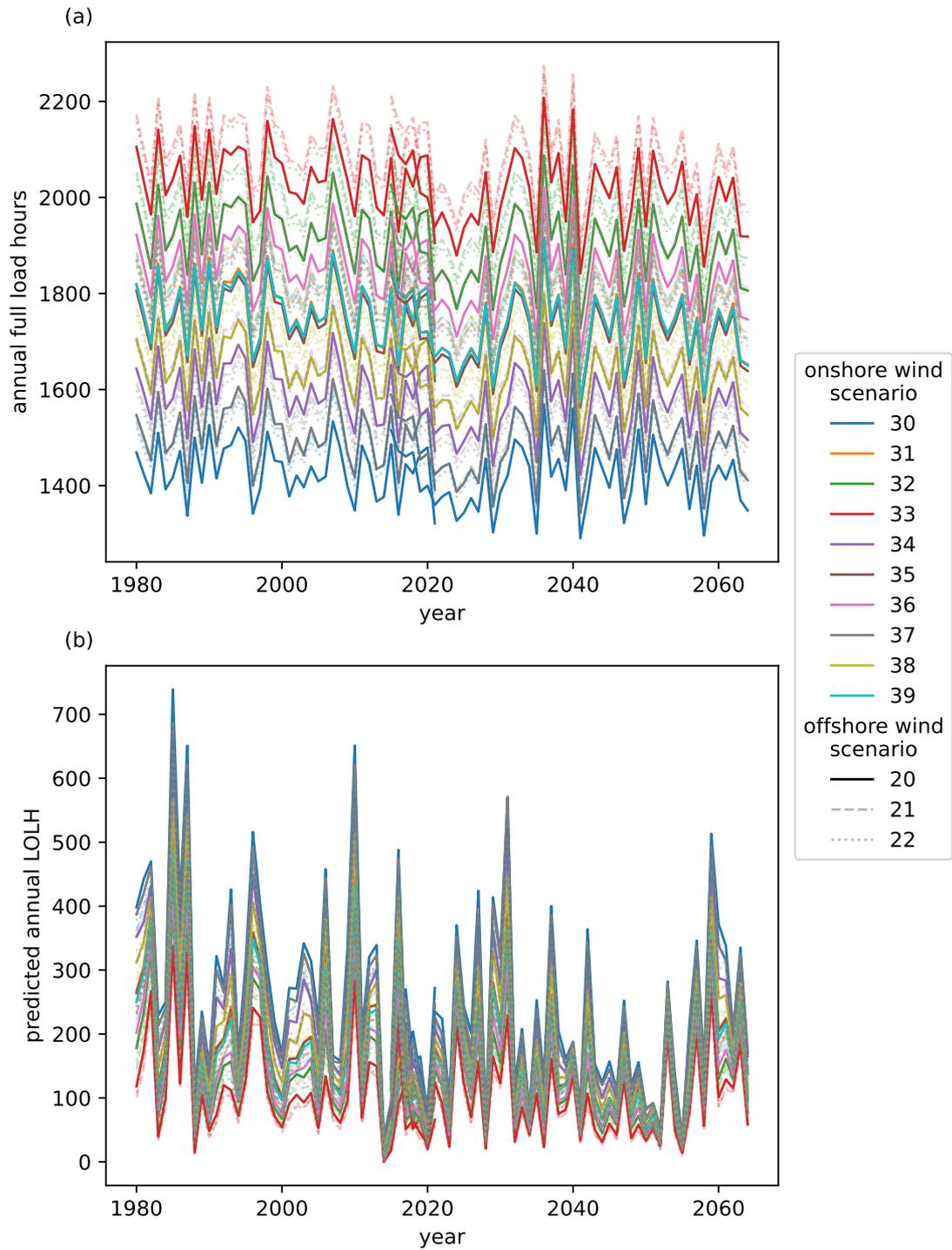


Figure 20: Annual full load hours (Panel (a)) and predicted annual LOLH (Panel (b)) for the 30 combinations of the 10 technical onshore wind scenarios (distinguished by colour) and 3 technical offshore wind scenarios (distinguished by line style) of the PECD 4.1. Results are shown for the ERA5 based reanalysis data and the MEHR climate model.

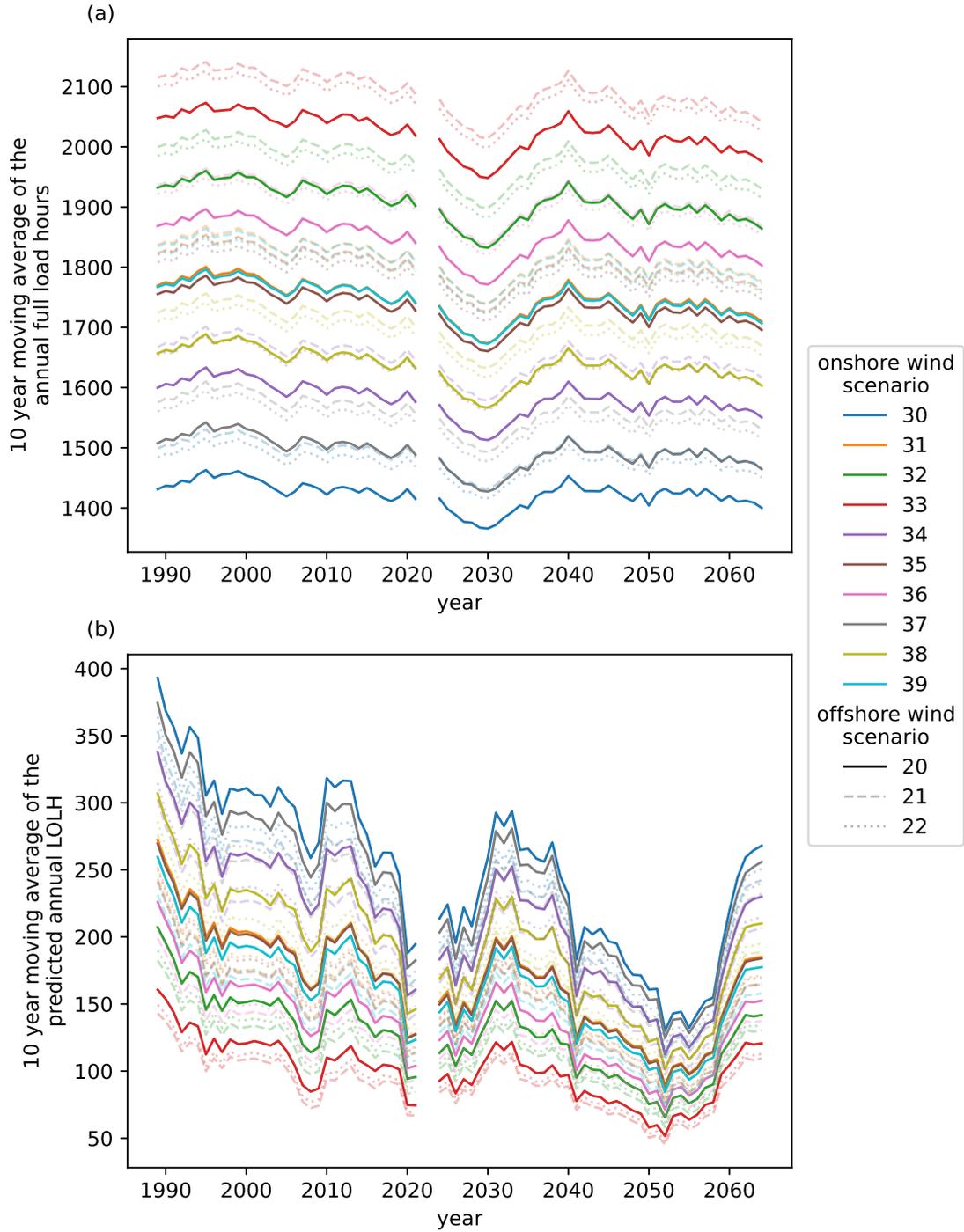


Figure 21: 10 year running mean of annual full load hours (panel (a)) and predicted annual LOLH (panel (b)) for the 30 combinations of the 10 technical onshore wind scenarios (distinguished by colour) and 3 technical offshore wind scenarios (distinguished by line style) of the PECD 4.1. Results are shown for the reanalysis data and the MEHR climate model.

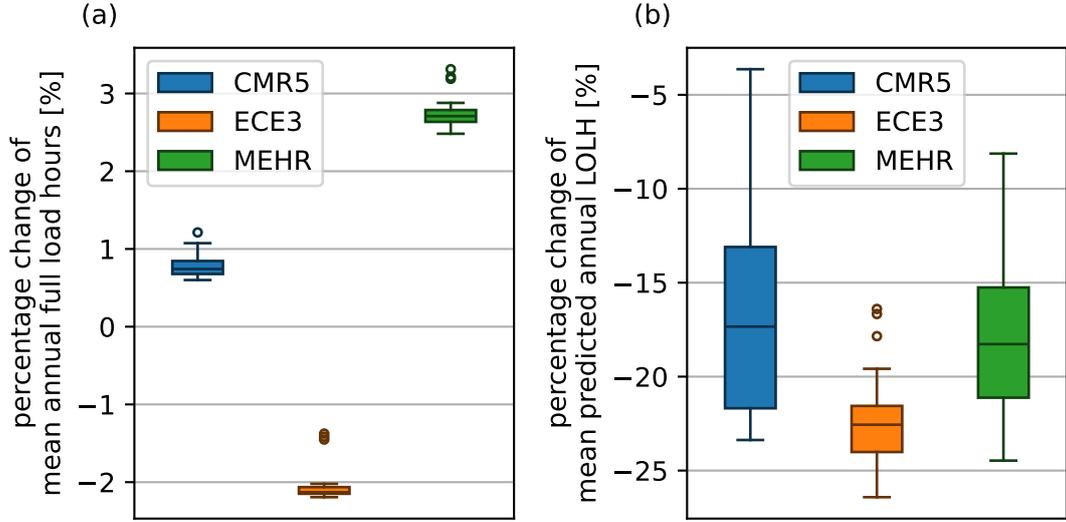


Figure 22: Box plots of the percentage change of the mean annual full load hours (panel (a)) and the mean predicted annual LOLH for the change from the reanalysis based dataset to the climate projections in the PECD 4.1. Hereby, the mean is taken over the years of the datasets. Each box corresponds to one of the three considered climate models and shows the distribution over the 30 technical scenarios.

LOLH, $\Delta\% \overline{\text{LOLH}}_{\text{model,on,off}}$, is given by

$$\Delta\% \overline{\text{LOLH}}_{\text{model,on,off}} = \frac{\sum_{\text{year}=2015}^{2064} \text{LOLH}_{\text{model,on,off}}^{\text{year}} - \sum_{\text{year}=1980}^{2021} \text{LOLH}_{\text{ERA5,on,off}}^{\text{year}}}{\sum_{\text{year}=1980}^{2021} \text{LOLH}_{\text{ERA5,on,off}}^{\text{year}}} \times 100. \quad (26)$$

The equation for the percentage change of the mean annual full load hours looks analogously. Figure 22 shows box plots of the distribution over the 30 technical scenarios of this percentage change for each of the three climate models. The change in the full load hours is not very large and actually the CMR5 and MEHR model gain full load hours compared to the historical data while the ECE3 model loses about 2.13% in the median. The change in the predicted LOLH is much more pronounced and all three models predict a significant decrease of 17.34% to 22.56% in the median. Interestingly, the ECE3 model shows the biggest decrease

although it had a decrease in full load hours. Apparently, the negative correlation between a change in full load hours and LOLH is not very strong. Considering the earlier observations of this thesis, this might not come as a surprise, as the full load hours do not incorporate the demand side of the electricity balance that is so crucial for the occurrence of electricity shortages.

A look at the annual mean temperatures, depicted in Figure 23, gives an idea why the projections predict so significantly less LOLH. The temperature in the PECD 4.1 was not detrended as was done for earlier versions of the ERAA as this was only a measure to obtain climate data that was closer aligned with estimations for the target year. Consequently, the temperature has a significant increasing trend that causes my indicator to predict less electricity shortages the further the years progress. The box plots in panel (b) of Figure 23 also give a plausible explanation why the ECE3 model showed such an increase in annual LOLH compared to the reanalysis data although the full load hours were lower. It has the highest temperatures of the three climate models and a median temperature that is about 1.50°C above the one of the historical dataset and maybe more importantly a high minimum annual mean temperature at about 8.50°C . Of course the annual mean temperatures can also only give a hint as the more important figure for adequacy concerns are the temperature extremes.

However, the original goal of my analysis was to assess the impact of the change from reanalysis data to climate projections in the ERAA 2024, and the earlier versions of the ERAA used detrended temperatures that were adjusted to align with temperatures expected for the target year. Therefore, a more appropriate test might be to compare the results of the reanalysis data with such a detrending to the results of the projections for the 12 weather years from 2025 to 2036 used in the ERAA 2024.

7.1.2. Analysis of the differences between temperature adjusted reanalysis data and climate projections in the PECD 4.1

In order to obtain temperatures that are comparable to the ones expected in the target year 2030, I detrended the temperature time series following a methodology described in Annex 2 of the ERAA 2023 report (ENTSO-E, 2025a, pp. 56–57).

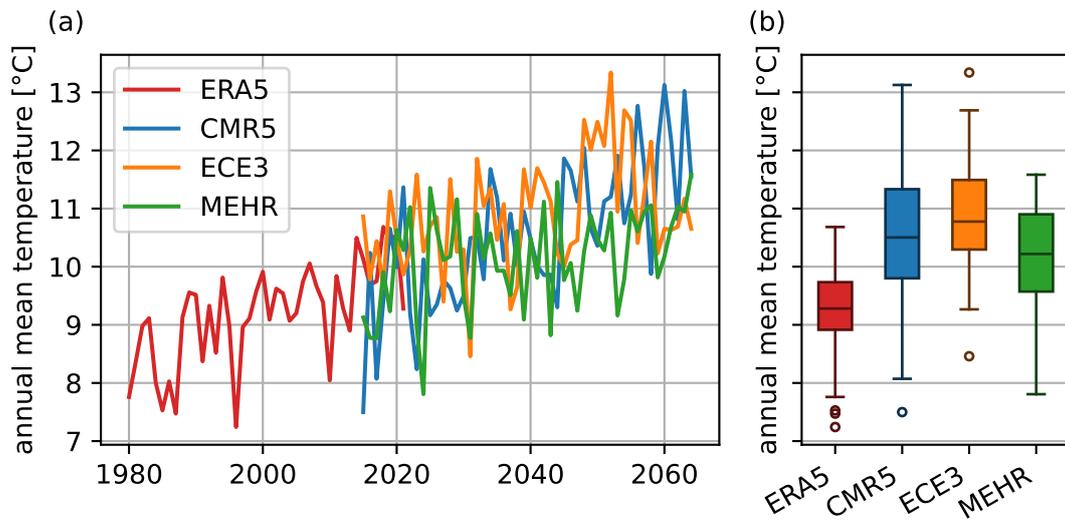


Figure 23: Line plot (panel (a)) and box plots (panel (b)) of annual mean temperatures for the ERA5 based historical dataset and the three climate models of the PECD 4.1. The box plots characterize the distribution over the years of the corresponding dataset.

Specifically, I chose the second methodology, detrending each month individually, combined with the first extrapolation adjustment approach, adjusting only the mean and not the standard deviation of the temperature. The details of the detrending are described in the following paragraph.

First, monthly mean temperatures are computed. Then, for each month a linear trend over the years is determined using a linear regression model. The slope of these regressions can be understood as the temperature adjustment needed per year of the difference between the target year and the year of the month that is to be adjusted. To avoid jumps between months that are adjusted by different amounts, the slopes are interpolated to an hourly time series over the whole year using the Piecewise Cubic Hermite Interpolating Polynomial method (PCHIP) (Fritsch & Carlson, 1980). The PCHIP method provides a rather smooth interpolation compared to a linear approach without overshooting extrema like a general spline approach might.⁷

⁷More precisely, the interpolating function is continuously differentiable and monotone between data points (Fritsch & Carlson, 1980).

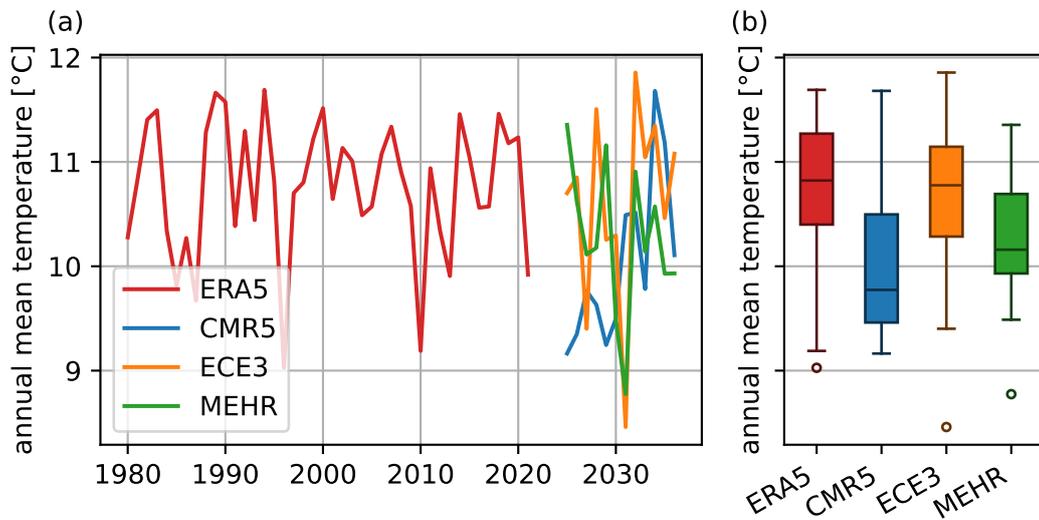


Figure 24: Line plot (panel (a)) and box plots (panel (b)) of annual mean temperatures. The ERA5 based historical data of the PECD 4.1 was adjusted to match the climate expected for the target year 2030 by following a linear trend. The projections are restricted to the 12 years used in the ERAA 2024. The box plots characterise the distribution over the years of the corresponding dataset.

Figure 24 shows the results of the detrending together with the 12 weather years of projections considered in the ERAA 2025. The box plot in Panel (b) shows, that the temperatures of the reanalysis data are now comparable to the ones of the ECE3 model which also has the highest temperatures of the climate projections in the considered subset of weather years. The CMR5 has the lowest median temperature but, on the other hand, the highest minimum.

The effects of the temperature adjustment and the restriction of weather years on the annual full load hours and the predicted annual LOLH can be seen in Figure 25. Analogously to Figure 22, it shows the distributions over the 30 technical scenarios of the percentage change of the 2 variables between the reanalysis dataset with adjusted temperatures and the climate projections. Panel (a) looks very similar to its counterpart in Figure 22, although the values for the CMR5 and MEHR model are a bit lower. Apparently, the change in mean full load hours is similar for the subset of weather years as on the whole set. The percentage change for the mean

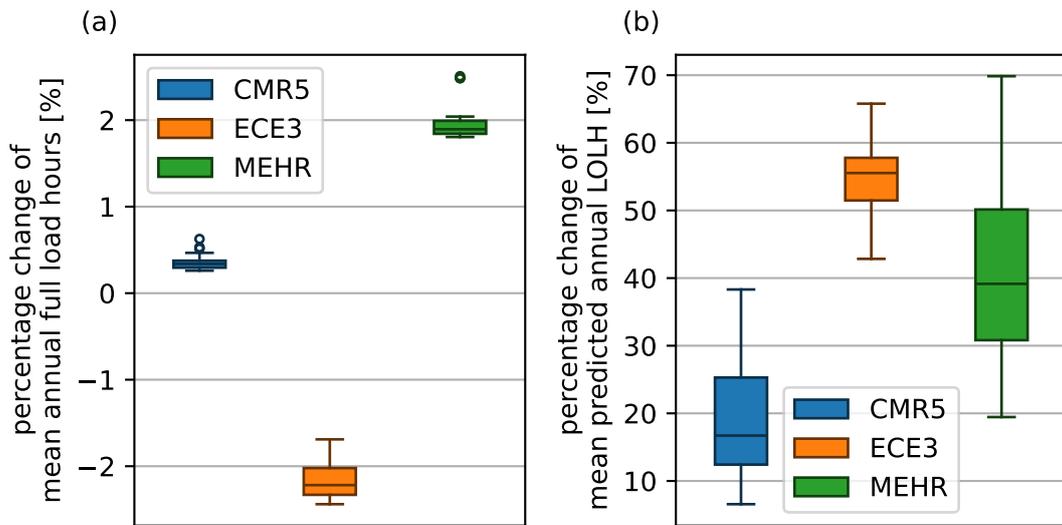


Figure 25: Box plots of the percentage change of the mean annual full load hours (panel (a)) and the mean predicted annual LOLH (panel (b)) for the change from the temperature adjusted reanalysis dataset to the 12 years of climate projections used in the ERAA 2024. Hereby, the mean is taken over the years of the datasets. Each box corresponds to one of three considered climate models and shows the distribution over the 30 technical scenarios of the PECD 4.1.

of the predicted annual LOLH is very different though to the one observed without the temperature adjustment. Now, the predicted LOLH increase significantly for all three climate models with a median change over all 90 scenarios of about 39.14%. The increase for the ECE3 model is actually the highest in the median with a value of 55.53%, although it has the highest median temperature. Part of the explanation is probably given by its low full load hours and the one outlier year with the lowest mean temperature in the comparison. However, I assume that an analysis of the corresponding time series with a higher resolution is necessary to get a satisfactory answer.

7.2. Evaluation

The analysis of the different technical scenarios showed, that the onshore scenarios 31, 35 and 39 show very similar results. This offers an opportunity to save com-

putational resources by only considering one of these. The significantly increased LOLH predictions for the climate projections shows that the choice of climate data from either reanalysis or projections is important. Interestingly, the power flow simulations predict substantially less electricity shortages for the ERAA 2024 with data from projections than for the ERAA 2022 with reanalysis based data. Seeing the even greater impact of the technical scenarios on the predicted LOLH, compared to the choice of climate data, this can likely be explained by the simultaneous changes in technical assumptions between the two ERAAs. In conclusion, I suggest the development of a realistic scenario for the spatial distribution and technical specifications of onshore wind generators should be a priority in the preparations for future ERAAs.

8. Conclusion

The reliable evaluation of resource adequacy concerns is a central challenge in the transition towards a climate neutral energy system. However, the amount of data that needs to be processed to factor in uncertainties such as climate variability can easily overwhelm the capacities of elaborate models. Therefore, a simplified approach that allows a preselection of the most relevant subsets of data can be of great value. I evaluated five methods to detect electricity shortages that are able to process thousands of scenarios within minutes. Commonly used VRE drought or dark doldrum indicators, that focus on the supply side of the electricity balance, performed poorly and should not be used for the task. In accordance with the results of Biewald et al. (2025), I found that the incorporation of the demand side is essential for a reliable detection of electricity shortages. The best results are achieved by the straightforward approach based on the residual load after Otero et al. (2022). However, its reliance on demand data, which is not always available, can be a hurdle for its application. The data-driven indicator I developed serves as an alternative, as it only uses temperature and time data as proxies for the electricity demand, and it performs almost as well as the residual load approach if sufficient training data is available. While both indicators need data to calibrate their parameters, the residual load method only has one parameter in contrast to the 96 of my approach. Consequently, the residual load approach

is less sensitive to a reduction of the training data. Comparing the results to the ones of Biewald et al. (2025), it is evident that an hourly instead of a daily temporal resolution substantially improves the F_1 -score. Furthermore, for most applications, the F_1 -score was found to be preferable to the F_2 -score as objective function in the optimisation.

Overall, the two indicators combining the supply and demand side are able to detect electricity shortages sufficiently well to allow for the reliable preselection of the most critical weather years and the identification of similar technical scenarios. Thus, they are appropriate tools to reduce the input data of resource adequacy assessments, allowing to account for a great range of uncertainties while keeping the amount of data that needs to be processed by computationally demanding simulations manageable. This can not only save money in the making of adequacy assessments, but, by helping to ensure that the available resources are used on relevant data, it can also improve the reliability of the results of the assessment (Biewald et al., 2025). This in return has the potential to reduce the costs of the overall power system, for example, by aiding the implementation of appropriately scaled capacity mechanisms.

Beyond their utility for resource adequacy assessments, the indicators are a valuable toolset for researchers investigating other sorts of questions involving electricity shortages, especially if power flow simulations are not feasible. In this context, my indicator can be particularly useful, as climate data is often more readily available than demand estimations. This was demonstrated in an analysis of the impact of the change from temperature adjusted reanalysis based climate data to climate projections in the ERAA 2024. I found that the change of data source leads to increased adequacy concerns, with mean annual LOLH increasing by 39.14% in the median of the considered scenarios. While this is a substantial change, switching from one technical scenario for the onshore wind production to another can have an even larger impact. Consequently, devising a well designed technical scenario for the ERAAs might be even more important than the choice of the best climate models and CO₂-reduction pathways.

Summarising, my analysis supports earlier findings by Biewald et al. (2025) that suggest that computationally efficient methods for the detection of electricity shortages can greatly benefit the study of resource adequacy concerns if they

incorporate the demand and supply side of the electricity balance. As my main contribution, I developed a novel SVM-based method that, given sufficient training data, provides reliable results without depending on demand data.

Acronyms

AC Air Conditioner. 58

ACER European Union Agency for the Cooperation of Energy Regulators. 6, 7, 93–96

C3S Copernicus Climate Change Service. 19, 80

CBT Constantly-Below-Threshold. 14, 15, 17, 55, *See Section 2.6.1.*

CMR5 CMCC-CM2-SR5. 22, 69, 72, *Glossary: CMR5*

DLR German Aerospace Center. 18, 23, 24

DSC Dice similarity coefficient. 12

DSO distribution system operator. 19, 94

DTU Technical University of Denmark. 19–21

ECE3 EC-EARTH3. 22, 69, 70, 72, 73, *Glossary: ECE3*

ECMWF European Centre for Medium-Range Weather Forecasts. 19, 22, 80

ED economic dispatch. 18, 24, 96

EENS expected energy not served. 96

ENS energy not served. 24

ENTSO-E European Network of Transmission System Operators for Electricity. 1, 6, 19, 24, 34, 58, 93–95

ERAA European Resource Adequacy Assessment. 1, 2, 6, 7, 14, 18–20, 22, 24, 30–32, 34, 35, 37–43, 45–53, 57–66, 70, 72–75, 93, 95, 96

ETSO European Transmission System Operators. 93

EU European Union. 6, 7, 19, 24, 93–95

EVA economic viability assessment. 6, 20, 96

FMBT fixed-duration mean below threshold. 15–17, 30, 34, 36, 37, 48, 54, *See Section 2.6.2*

IPCC Intergovernmental Panel on Climate Change . 2, 80, *Glossary: IPCC*

LOLE loss of load expectation. 96

LOLH loss of load hours. 5, 8, 9, 12, 17, 18, 29, 32, 34–37, 40–42, 45–49, 51–54, 63, 66–70, 72–75

MAD mean absolute deviation. 10, 11, 35

MAE mean absolute error. 10, 11, 34–36, 39, 40, 45–47, 51, 52, 60–63

MASE mean absolute scaled error. 10, 11, 34–36, 39, 40, 45–47, 51, 52, 63

MBT Mean-Below-Threshold. 15, 16, *See Section 2.6.2*

MEHR MPI-ESM1-2-HR. 22, 66–69, 72, *Glossary: MEHR*

NECP National Energy and Climate Plan. 19, 96

NRA national regulatory agency. 94

PECD Pan-European Climate Database. 19–23, 25, 33, 64, 65, 67–73

PEMMDB Pan-European Market Modelling Database. 19, 20, 65

PRL positive residual load. 17

PV photovoltaics. 1, 3, 4, 14, 20, 22, 42, 55, 63, 65

RMSE root mean square error. 10

SPA Sequent Peak Algorithm. 16, 17, 30, 34, 37

SSP Shared Socioeconomic Pathway. 2

SVM support vector machine. 25, 26, 28, 29, 44, 46, 47, 57, 62–64, 76

TRF Temperature Response Function. 58, 60–62

TSO transmission system operator. 6, 19, 22–24, 58, 93, 94, 96

UN United Nations. 80

VMBT variable-duration mean below threshold. 15–17, *See Section 2.6.2*

VRE variable renewable energy. ii, 3, 4, 13–18, 20, 25, 30, 32–34, 36, 37, 42, 45, 48, 50, 54, 55, 74

Glossary

CMIP6 Stands for Coupled Model Intercomparison Project Phase 6. It is a project to coordinate a common framework for different climate models to allow for intercomparison. The Models are used by the IPCC. 19, 80

CMR5 Is short for CMCC-CM2-SR5. It is an earth system model developed at the Euro-Mediterranean Center on Climate Change in Bologna, Italy, that was used for the CMIP6 (Lovato & Peano, 2020). 22, 77

ECE3 Is short for EC-EARTH3. It is an earth system model developed by the EC-EARTH consortium, version 3.3 of which was used for the CMIP6 (Döscher et al., 2022). 22, 77

ERA5 Is the fifth generation of atmospheric reanalysis created by C3S at ECMWF. 19–21, 67, 71, 72

IPCC Stands for Intergovernmental Panel on Climate Change and is the body of the United Nations (UN) that assesses the science related to climate change.. 2, 78

MEHR Is short for MPI-ESM1.2-HR. It is an earth system model of the Max Planck Institute that was used for the CMIP6 (von Storch et al., 2017). 22, 78

PyPSA Stands for Python for Power System Analysis. It is a project providing a comprehensive open-source python environment for power system modeling. 24

References

- ACER. (2020, February 10). Methodology for the european resource adequacy assessment - ACER decision on the ERAA methodology: Annex i. Retrieved February 7, 2025, from https://www.acer.europa.eu/sites/default/files/documents/Individual%20Decisions_annex/ACER%20Decision%2024-2020%20on%20ERAA%20-%20Annex%20I_1.pdf
- ACER. (2022, February 22). DECISION no 02/2022 OF THE EUROPEAN UNION AGENCY FOR THE COOPERATION OF ENERGY REGULATORS of 22 february 2022 on the european resource adequacy assessment for 2021. Retrieved June 13, 2025, from https://www.acer.europa.eu/sites/default/files/documents/Individual%20Decisions/ACER%20Decision%2002-2022%20on%20ERAA%202021_0.pdf
- ACER. (2023, February 27). DECISION no 04/2023 OF THE EUROPEAN UNION AGENCY FOR THE COOPERATION OF ENERGY REGULATORS of 27 february 2023 on the european resource adequacy assessment for 2022. Retrieved February 4, 2025, from https://www.acer.europa.eu/sites/default/files/documents/Individual%20Decisions/ACER_Ddecision_04-2023_ERAA_2022.pdf
- Avdijaj, B., Henneaux, P., & Labeau, P.-E. (2024). Alternative adequacy indicators to complement the LOLE in power systems with a large share of renewable energy sources. *2024 International Conference on Renewable Energies and Smart Technologies (REST)*, 1–5. <https://doi.org/10.1109/REST59987.2024.10645402>
- Biewald, B., Cozian, B., Dubus, L., Zappa, W., & Stoop, L. P. (2025). Evaluation of ‘dunkelflaute’ event detection methods considering grid operators’ needs [Publisher: IOP Publishing]. *Environmental Research: Energy*, 2(2), 025007. <https://doi.org/10.1088/2753-3751/adcf29>
- Bloomfield, H. C., Gonzalez, P. L. M., Lundquist, J. K., Stoop, L. P., Browell, J., Dargaville, R., Felice, M. D., Gruber, K., Hilbers, A., Kies, A., Panteli, M., Thornton, H. E., Wohland, J., Zeyringer, M., & Brayshaw, D. J. (2021). The importance of weather and climate to energy systems: A workshop on next generation challenges in energy–climate modeling [Publisher:

- American Meteorological Society Section: Bulletin of the American Meteorological Society]. *Bulletin of the American Meteorological Society*, 102(1), E159–E167. <https://doi.org/10.1175/BAMS-D-20-0256.1>
- Bloomfield, H., & Brayshaw, D. (2021). ERA5 derived time series of european aggregated surface weather variables, wind power, and solar power capacity factors: Hourly data from 1950-2020. <https://doi.org/10.17864/1947.000321>
- Bloomfield, H. C., Brayshaw, D. J., Deakin, M., & Greenwood, D. (2022). Hourly historical and near-future weather and climate variables for energy system modelling [Publisher: Copernicus GmbH]. *Earth System Science Data*, 14(6), 2749–2766. <https://doi.org/10.5194/essd-14-2749-2022>
- Brown, T., Hörsch, J., & Schlachtberger, D. (2018). PyPSA: Python for Power System Analysis. *Journal of Open Research Software*, 6(4). <https://doi.org/10.5334/jors.188>
- Bundesnetzagentur. (2025). *Dunkelflaute? kein grund zur panik [dark doldrums? no reason to panic]* [Insight blog]. Retrieved June 11, 2025, from <https://www.bundesnetzagentur.de/DE/Allgemeines/DieBundesnetzagentur/Insight/Texte/Energiewende/Dunkelflaute.html>
- Burger, B. (n.d.). *Energy-charts*. Retrieved June 11, 2025, from <https://www.energy-charts.info>
- Capacity factor* [NRC web]. (2023, February 15). Retrieved March 21, 2025, from <https://www.nrc.gov/reading-rm/basic-ref/glossary/capacity-factor.html>
- Castillo, V. Z., Boer, H.-S. d., Muñoz, R. M., Gernaat, D. E. H. J., Benders, R., & van Vuuren, D. (2022). Future global electricity demand load curves. *Energy*, 258, 124741. <https://doi.org/10.1016/j.energy.2022.124741>
- Christen, P., Hand, D. J., & Kirielle, N. (2023). A review of the f-measure: Its history, properties, criticism, and alternatives. *ACM Comput. Surv.*, 56(3), 73:1–73:24. <https://doi.org/10.1145/3606367>
- Connolly, D. (2017). Heat roadmap europe: Quantitative comparison between the electricity, heating, and cooling sectors for different european countries. *Energy*, 139, 580–593. <https://doi.org/10.1016/j.energy.2017.07.037>
- Copernicus Climate Change Service. (2019). ERA5 monthly averaged data on pressure levels from 1940 to present. <https://doi.org/10.24381/CDS.6860A573>

- Copernicus Climate Change Service (C3S) Climate Data Store (CDS). (2024). Climate and energy related variables from the pan-european climate database derived from reanalysis and climate projections. <https://doi.org/10.24381/CDS.F323C5EC>
- Deutschlandfunk. (2025, January 7). *Dunkelflaute und energiewende: Wie sicher ist deutschlands strom? [dark doldrums and the energy transition: How safe is germany's electricity?]* [Deutschlandfunk]. Retrieved June 11, 2025, from <https://www.deutschlandfunk.de/strom-deutschland-strompreis-erneuerbare-energien-dekarbonisierung-100.html>
- Dice, L. R. (1945). Measures of the amount of ecologic association between species [Publisher: [Wiley, Ecological Society of America]]. *Ecology*, *26*(3), 297–302. <https://doi.org/10.2307/1932409>
- Do, L. P. C., Hagfors, L. I., Lin, K.-H., & Molnár, P. (2016). Demand and residual demand modelling using quantile regression [Publisher: EDP Sciences]. *Renewable Energy and Environmental Sustainability*, *1*, 41. <https://doi.org/10.1051/rees/2016045>
- Dodge, Y. (2010). *The concise encyclopedia of statistics*. Springer.
- Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arsouze, T., Bergman, T., Bernardello, R., Boussetta, S., Caron, L.-P., Carver, G., Castrillo, M., Catalano, F., Cvijanovic, I., Davini, P., Dekker, E., Doblas-Reyes, F. J., Docquier, D., Echevarria, P., Fladrich, U., ... Zhang, Q. (2022). The EC-earth3 earth system model for the coupled model intercomparison project 6 [Publisher: Copernicus GmbH]. *Geoscientific Model Development*, *15*(7), 2973–3020. <https://doi.org/10.5194/gmd-15-2973-2022>
- du Toit, S. H. C., Steyn, A. G. W., & Stumpf, R. H. (1986). *Graphical exploratory data analysis*. Springer. <https://doi.org/10.1007/978-1-4612-4950-4>
- ECMWF. (n.d.). *Climate and energy related variables from the pan-european climate database derived from reanalysis and climate projections: Product user guide (PUG)*. Retrieved May 29, 2025, from <https://confluence.ecmwf.int/x/a78zGg>
- ENTSO-E. (n.d.-a). ENTSO-e mission statement. Retrieved February 21, 2025, from <https://eepublicdownloads.azureedge.net/clean-documents/General>

- [entso-e_mission_statement_211231_Final.pdf](https://www.entsoe.eu/eraa/2022/downloads/entso-e_mission_statement_211231_Final.pdf)
- ENTSO-E. (n.d.-b). *ERAA 2022 Downloads*. Retrieved May 28, 2025, from <https://www.entsoe.eu/eraa/2022/downloads/>
- ENTSO-E. (n.d.-c). *ERAA 2024 downloads*. Retrieved May 28, 2025, from <https://www.entsoe.eu/eraa/2024/downloads/>
- ENTSO-E. (2022). European resource adequacy assessment - 2022 edition - annex 1 – input data & assumptions. Retrieved June 10, 2025, from https://eepublicdownloads.blob.core.windows.net/public-cdn-container/clean-documents/sdc-documents/ERAA/2022/data-for-publication/ERAA2022_Annex_1_Assumptions.pdf
- ENTSO-E. (2024a, July 24). European resource adequacy assessment 2024 - data collection guidelines - draft v0.1. Retrieved June 15, 2025, from https://eepublicdownloads.blob.core.windows.net/public-cdn-container/clean-documents/sdc-documents/ERAA/2024/ERAA_2024_-_Data_collection_guidelines.pdf
- ENTSO-E. (2024b, September 4). TYNDP 2024 - opportunities for a more efficient european power system by 2050 - infrastructure gaps report. Retrieved June 10, 2025, from https://eepublicdownloads.blob.core.windows.net/public-cdn-container/tyndp-documents/TYNDP2024/foropinion/Infrastructure_Gaps_Report.pdf
- ENTSO-E. (2025a, February). European Resource Adequacy Assessment 2023 Edition - Annex 2: Methodology - Corrigendum 18/02/2025. Retrieved May 31, 2025, from https://www.entsoe.eu/eraa/2023/report/ERAA_2023_Annex_2_Methodology.pdf
- ENTSO-E. (2025b, April). European Resource Adequacy Assessment - 2024 Edition - Executive Report. Retrieved June 12, 2025, from https://eepublicdownloads.blob.core.windows.net/public-cdn-container/clean-documents/sdc-documents/ERAA/2024/report/ERAA_2024_Executive_Report.pdf
- ENTSO-E. (2025c, April). European resource adequacy assessment - 2024 edition - executive report. Retrieved June 12, 2025, from https://eepublicdownloads.blob.core.windows.net/public-cdn-container/clean-documents/sdc-documents/ERAA/2024/report/ERAA_2024_Executive_Report.pdf

- Etling, D. (2008). Einführung in die numerische wettervorhersage und klimamodellierung [introduction to numerical weather prediction and climate modeling]. In *Theoretische meteorologie: Eine einföhrung* (pp. 243–254). Springer. https://doi.org/10.1007/978-3-540-75979-9_17
- European Commission. (n.d.). *Transparency register - organisation detail*. Retrieved June 5, 2025, from https://transparency-register.europa.eu/searchregister-or-update/organisation-detail_en
- Fritsch, F. N., & Carlson, R. E. (1980). Monotone piecewise cubic interpolation [Publisher: Society for Industrial and Applied Mathematics]. *SIAM Journal on Numerical Analysis*, 17(2), 238–246. <https://doi.org/10.1137/0717021>
- Gerke, T. (2014, August 5). *Energy transition 101 - energy transition vocabulary: Residual load* [Section: Speicher]. Retrieved March 21, 2025, from <https://www.erneuerbareenergien.de/transformation/speicher/energy-transition-101-energy-transition-vocabulary-residual-load>
- Global wind atlas*. (n.d.). Retrieved May 28, 2025, from <https://globalwindatlas.info>
- Hagspiel, S., Knaut, A., & Peter, J. (2018). Reliability in multi-regional power systems: Capacity adequacy and the role of interconnectors [Publisher: SAGE Publications]. *The Energy Journal*, 39(5), 183–204. <https://doi.org/10.5547/01956574.39.5.shag>
- Han, J., Kamber, M., & Pei, J. (2012, January 1). 3 - data preprocessing. In J. Han, M. Kamber & J. Pei (Eds.), *Data mining: Concepts and techniques (third edition)* (pp. 83–124). Morgan Kaufmann. <https://doi.org/10.1016/B978-0-12-381479-1.00003-4>
- He, H. (2013). *Imbalanced learning: Foundations, algorithms, and applications* (1st ed). John Wiley & Sons, Incorporated.
- Heylen, E., Deconinck, G., & Van Hertem, D. (2018). Review and classification of reliability indicators for power systems with a high share of renewable energy sources. *Renewable and Sustainable Energy Reviews*, 97, 554–568. <https://doi.org/10.1016/j.rser.2018.08.032>
- Ho, C. K., Roesler, E. L., Nguyen, T., & Ellison, J. (2023). Potential impacts of climate change on renewable energy and storage requirements for grid

- reliability and resource adequacy. *Journal of Energy Resources Technology*, 145(10), 100904. <https://doi.org/10.1115/1.4062891>
- Hu, W., Scholz, Y., Yeligeti, M., Deng, Y., & Jochem, P. (2024). Future electricity demand for europe: Unraveling the dynamics of the temperature response function. *Applied Energy*, 368, 123387. <https://doi.org/10.1016/j.apenergy.2024.123387>
- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688. <https://doi.org/10.1016/j.ijforecast.2006.03.001>
- Hyndman (<https://stats.stackexchange.com/users/159/rob-hyndman>), R. (2014). Alternative to mape when the data is not a time series [version: 2014-07-23]. Retrieved June 20, 2025, from <https://stats.stackexchange.com/q/108963>
- Ibanez, E., & Milligan, M. (2014). Comparing resource adequacy metrics and their influence on capacity value. *2014 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, 1–6. <https://doi.org/10.1109/PMAPS.2014.6960610>
- Intergovernmental Panel On Climate Change (Ippc). (2023, July 6). *Climate change 2021 – the physical science basis: Working group i contribution to the sixth assessment report of the intergovernmental panel on climate change* (1st ed.). Cambridge University Press. <https://doi.org/10.1017/9781009157896>
- Janal, Y., Regner, P., & Schmidt, J. (2025). More wind, falling efficiency: The driving factors of german onshore wind power generation. *Wind Energy*, 28(4), e70013. <https://doi.org/10.1002/we.70013>
- Janzing, B. (2024). Ungelöstes problem der erneuerbaren: Ein november voller dunkelflauten [unsolved problem of renewables: A november full of dark doldrums]. *Die Tageszeitung: taz*. Retrieved June 11, 2025, from <https://taz.de/Ungeloestes-Problem-der-Erneuerbaren/!6045762/>
- Jeppesen, J. (2023, August 23). *Fact sheet: Reanalysis* [ECMWF]. Retrieved June 5, 2025, from <https://www.ecmwf.int/en/about/media-centre/focus/2023/fact-sheet-reanalysis>
- Johansson, V., Thorson, L., Goop, J., Göransson, L., Odenberger, M., Reichenberg, L., Taljegard, M., & Johnsson, F. (2017). Value of wind power – implications

- from specific power. *Energy*, 126, 352–360. <https://doi.org/10.1016/j.energy.2017.03.038>
- Kapica, J., Jurasz, J., Canales, F. A., Bloomfield, H., Guezgouz, M., De Felice, M., & Kobus, Z. (2024). The potential impact of climate change on european renewable energy droughts. *Renewable and Sustainable Energy Reviews*, 189, 114011. <https://doi.org/10.1016/j.rser.2023.114011>
- Kaspar, F., Bär, Franziska, Drücke, Jaqueline, James, Paul, Ostermöller, Jennifer & Zepperitz, Magdalena. (2024, December 17). *Klimatologische einordnung der „dunkelflaute“ im november 2024 [climatological classification of the "dark doldrums" in november 2024]*. Deutscher Wetterdienst.
- Kittel, M., & Schill, W.-P. (2024, August 1). Measuring the dunkelflaute: How (not) to analyze variable renewable energy shortage. <https://doi.org/10.48550/arXiv.2402.06758>
- Klein, O. (2024, December 13). *Dunkelflaute lässt strompreise explodieren [dark doldrums cause electricity prices to explode]* [ZDFheute]. Retrieved June 11, 2025, from <https://www.zdfheute.de/wirtschaft/strompreise-rekord-dunkelflaute-100.html>
- Koivisto, M., Kanellas, P., Troccoli, A., Aldrigo, G., Amaro e Silva, R., Olsen, B. T., Murcia, J. P., Angeloni, D., Borga, M., Campostrini, S., Cordeddu, S., Lusito, L., Saint-Drenan, Y. .-, Restivo, E., & Zaramella, M. (2023). Developing support service to ENTSO-e: Including the impacts of climate change in the pan-european climate database (PECD). *22nd Wind and Solar Integration Workshop (WIW 2023)*, 2023, 58–66. <https://doi.org/10.1049/icp.2023.2719>
- Koivisto, M. J., & Murcia Leon, J. P. (2022, January 1). Pan-european wind and solar generation time series (PECD 2021 update). <https://doi.org/10.11583/DTU.C.5939581.V3>
- Krapp, C. (2024). Strompreis erreicht höchsten wert seit der energiekrisis [electricity price reaches highest value since energy crisis]. *Handelsblatt*. Retrieved June 11, 2025, from <https://www.handelsblatt.com/unternehmen/energie/energie-strompreis-erreicht-hoechsten-wert-seit-der-energiekrise/100086236.html>

- Lovato, T., & Peano, D. (2020). CMCC CMCC-CM2-SR5 model output prepared for CMIP6 ScenarioMIP. <https://doi.org/10.22033/ESGF/CMIP6.1365>
- Luburić, Z., Pandžić, H., Plavšić, T., Teklić, L., & Valentić, V. (2018). Role of energy storage in ensuring transmission system adequacy and security. *Energy*, *156*, 229–239. <https://doi.org/10.1016/j.energy.2018.05.098>
- Lüers, S., & Heyken, M. (2024). Status of onshore wind energy development in germany - year 2024.
- Luo, J., Rohn, J., Bayer, M., Priess, A., Wilkmann, L., & Xiang, W. (2015). Heating and cooling performance analysis of a ground source heat pump system in southern germany. *Geothermics*, *53*, 57–66. <https://doi.org/10.1016/j.geothermics.2014.04.004>
- Mockert, F., Grams, C. M., Brown, T., & Neumann, F. (2023). Meteorological conditions during periods of low wind speed and insolation in germany: The role of weather regimes. *Meteorological Applications*, *30*(4), e2141. <https://doi.org/10.1002/met.2141>
- Murcia, J. P., Koivisto, M. J., Luzia, G., Olsen, B. T., Hahmann, A. N., Sørensen, P. E., & Als, M. (2022). Validation of european-scale simulated wind speed and wind generation time series. *Applied Energy*, *305*, 117794. <https://doi.org/10.1016/j.apenergy.2021.117794>
- Myers, J. L., & Well, A. D. (2003). *Research design and statistical analysis* (2. ed). Erlbaum.
- North American Electric Reliability Corporation. (2024, December 17). 2024 long-term reliability assessment. Retrieved March 31, 2025, from https://www.nerc.com/pa/RAPA/ra/Reliability%20Assessments%20DL/NERC_Long%20Term%20Reliability%20Assessment_2024.pdf
- Otero, N., Martius, O., Allen, S., Bloomfield, H., & Schaeffi, B. (2022). A copula-based assessment of renewable energy droughts across europe. *Renewable Energy*, *201*, 667–677. <https://doi.org/10.1016/j.renene.2022.10.091>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, *12*(85), 2825–2830. <http://jmlr.org/papers/v12/pedregosa11a.html>

- Pezzutto, S., Fazeli, R., De Felice, M., & Sparber, W. (2016). Future development of the air-conditioning market in Europe: An outlook until 2020. *WIREs Energy and Environment*, 5(6), 649–669. <https://doi.org/10.1002/wene.210>
- Pfenninger, S., & Staffell, I. (2016). Long-term patterns of european PV output using 30 years of validated hourly reanalysis and satellite data. *Energy*, 114, 1251–1265. <https://doi.org/10.1016/j.energy.2016.08.060>
- Potissomporn, P., Adcock, T. A. A., & Vogel, C. R. (2024). Extreme value analysis of wind droughts in great britain. *Renewable Energy*, 221, 119847. <https://doi.org/10.1016/j.renene.2023.119847>
- Regulation (EC) No 713/2009 of the European Parliament and of the Council of 13 July 2009 establishing an Agency for the Cooperation of Energy Regulators (2009, July 13). Retrieved June 5, 2025, from <http://data.europa.eu/eli/reg/2009/713/oj/eng>
- Regulation (EC) No 714/2009 of the European Parliament and of the Council of 13 July 2009 on conditions for access to the network for cross-border exchanges in electricity and repealing Regulation (EC) No 1228/2003 (2009, July 13). Retrieved June 5, 2025, from <http://data.europa.eu/eli/reg/2009/714/oj/eng>
- Regulation (EU) 2019/942 of the European Parliament and of the Council of 5 June 2019 establishing a European Union Agency for the Cooperation of Energy Regulators (recast) (2019, June 5). Retrieved June 5, 2025, from <http://data.europa.eu/eli/reg/2019/942/oj/eng>
- Regulation (EU) 2019/943 of the European Parliament and of the Council of 5 June 2019 on the internal market for electricity (recast) (2019, June 5). Retrieved June 5, 2025, from <http://data.europa.eu/eli/reg/2019/943/oj/eng>
- Rönsch, R. (2024). Kein wind, keine sonne: Was wir aus der dunkelflaute für die energiewende lernen [no wind, no sun: What we can learn from dark doldrums for the energy transition]. *MDR.DE*. Retrieved June 11, 2025, from <https://www.mdr.de/wissen/dunkelflaute-kein-wind-keine-sonne-auswirkungen-anforderungen-100.html>
- Saurugg, H. (2024). Dunkelflaute offenbart massive planungsfehler – was jetzt getan werden muss [dark doldrum reveals massive planning errors - what needs to be done now]. *FOCUS online*. Retrieved June 10, 2025, from <http://www.focus-online.de>

[s://www.focus.de/wissen/energie-spezialist-die-dunkelflaute-offenbart-massive-planungsfehler_id_260564773.html](https://www.focus.de/wissen/energie-spezialist-die-dunkelflaute-offenbart-massive-planungsfehler_id_260564773.html)

- Schwab, A. J. (2015). *Elektroenergiesysteme: Erzeugung, Übertragung und Verteilung elektrischer Energie [Electrical energy systems: generation, transmission and distribution of electrical energy]* (4. Aufl. 2015). Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-46856-2>
- Schyska, Bruno U., Witte, Francesco & Straub, Jonas. (2024, July). DestinE EnSys demonstrator [original-date: 2024-03-06T11:07:25Z]. Retrieved June 4, 2025, from https://github.com/dlr-ve-esy/DestinE_EnSys_Demonstrator
- Shen, C., Zhu, W., Tang, X., Du, W., Wang, Z., Xu, S., & Yao, K. (2024). Risk assessment and resilience enhancement strategies for urban power supply-demand imbalance affected by extreme weather: A case study of beijing. *International Journal of Disaster Risk Reduction*, *106*, 104471. <https://doi.org/10.1016/j.ijdr.2024.104471>
- Simoglou, C. K., & Biskas, P. N. (2023). Capacity mechanisms in europe and the US: A comparative analysis and a real-life application for greece [Number: 2 Publisher: Multidisciplinary Digital Publishing Institute]. *Energies*, *16*(2), 982. <https://doi.org/10.3390/en16020982>
- Slingo, J., & Palmer, T. (2011). Uncertainty in weather and climate prediction [Publisher: Royal Society]. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *369*(1956), 4751–4767. <https://doi.org/10.1098/rsta.2011.0161>
- Solar Power Europe. (2025, July 5). European market outlook for battery storage 2025-2029. Retrieved June 10, 2025, from https://api.solarpowereurope.org/uploads/SPE_European_Battery_Outlook_2025_62b89db476.pdf?updated_at=2025-05-05T09:16:53.946Z
- Spänhoff, B. (2014). Current status and future prospects of hydropower in saxony (germany) compared to trends in germany, the european union and the world. *Renewable and Sustainable Energy Reviews*, *30*, 518–525. <https://doi.org/10.1016/j.rser.2013.10.035>
- Staffell, I., & Pfenninger, S. (2016). Using bias-corrected reanalysis to simulate current and future wind power output. *Energy*, *114*, 1224–1239. <https://doi.org/10.1016/j.energy.2016.08.068>

- Tharwat, A. (2021). Classification assessment methods [Publisher: Emerald Publishing Limited]. *Applied Computing and Informatics*, 17(1), 168–192. <https://doi.org/10.1016/j.aci.2018.08.003>
- Tuning the decision threshold for class prediction* [Scikit-learn.org] [version 1.7.0]. (n.d.). Retrieved June 13, 2025, from https://scikit-learn/stable/modules/classification_threshold.html
- U.S. Energy Information Administration (EIA). (n.d.). Glossary - Generator nameplate capacity (installed). Retrieved June 12, 2025, from [https://www.eia.gov/tools/glossary/index.php?id=Generator%20nameplate%20capacity%20\(installed\)](https://www.eia.gov/tools/glossary/index.php?id=Generator%20nameplate%20capacity%20(installed))
- van der Wiel, K., Stoop, L. P., van Zuijlen, B. R. H., Blackport, R., van den Broek, M. A., & Selten, F. M. (2019). Meteorological conditions leading to extreme low variable renewable energy production and extreme high energy shortfall. *Renewable and Sustainable Energy Reviews*, 111, 261–275. <https://doi.org/10.1016/j.rser.2019.04.065>
- von Storch, J.-S., Putrasahan, D., Lohmann, K., Gutjahr, O., Jungclaus, J., Bittner, M., Haak, H., Wieners, K.-H., Giorgetta, M., Reick, C., Esch, M., Gayler, V., de Vrese, P., Raddatz, T., Mauritsen, T., Behrens, J., Brovkin, V., Claussen, M., Crueger, T., . . . Roeckner, E. (2017). MPI-m MPIESM1.2-HR model output prepared for CMIP6 HighResMIP. <https://doi.org/10.22033/ESGF/CMIP6.762>
- WDR. (2024). Dunkelflaute im November: Windstille und Dunkelheit ließen Strompreis steigen [Dark doldrums in November: Calm and darkness caused electricity prices to rise]. *WDR.DE*. Retrieved June 11, 2025, from <https://www1.wdr.de/nachrichten/dunkelflaute-strom-wind-solar-energie-deutschland-november-100.html>
- Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30(1), 79–82. <https://doi.org/10.3354/cr030079>
- Wind energy database*. (n.d.). Retrieved May 29, 2025, from <https://www.thewindpower.net/>

Zhou, Z.-H. (2021). *Machine learning*. Springer. <https://doi.org/10.1007/978-981-15-1967-3>

A. The European Resource Adequacy Assessment

In order to ensure a reliable electricity supply within a transforming energy system, a robust assessment of the adequacy of the available resources is of vital importance to allow policy makers and other stakeholders to come to informed decisions. Assessments on the national level are seen to not sufficiently recognize the situations in neighbouring countries.⁸ However, a coordinated approach to adequacy concerns is seen to be crucial to ensure properly functioning cross-border electricity markets, facilitating a more robust and cost effective system for all involved countries. Consequently, in 2019, the EU set the legal framework for the conduction of a European Resource Adequacy Assessment (ERAA). As the ERAA provides the context for which my work has the most relevance, I will introduce its legal foundations, involved parties and methodology before delving into the main part of my thesis.

A.1. Involved institutions

On the highest level, two European institutions are mainly responsible for the ERAA: the ENTSO-E for the creation and the ACER in a supervisory function.

A.1.1. The European Network of Transmission System Operators for Electricity (ENTSO-E):

The European Network of Transmission System Operators for Electricity (ENTSO-E) is the association of European TSOs, facilitating the cooperation of TSOs to ensure a reliable, coordinated operation of the European power grid. It was founded in 2009 as the successor of the European Transmission System Operators (ETSO).⁹ Membership in the ENTSO-E is mandatory for all TSOs in the EU.¹⁰ Its tasks include the development of operational standards and rules for the effective,

⁸Regulation (EU) 2019/942 of the European Parliament and of the Council of 5 June 2019 establishing a European Union Agency for the Cooperation of Energy Regulators (recast), 2019, Recital 5.

⁹Regulation (EC) No 713/2009 of the European Parliament and of the Council of 13 July 2009 establishing an Agency for the Cooperation of Energy Regulators, 2009.

¹⁰Regulation (EU) 2019/943 of the European Parliament and of the Council of 5 June 2019 on the internal market for electricity (recast), 2019, Article 28.

reliable and competitive operation of the European electricity market in the form of network codes. Furthermore, the ENTSO-E biennially publishes a Union-wide ten-year network development plan, supports the coordination of stakeholders and has to conduct resource adequacy assessments on different time scales. Beyond that the ENTSO-E sees itself as "the common voice of European TSOs" (ENTSO-E, n.d.-a) and claims to provide "expert contributions and a constructive view to energy debates to support policymakers in making informed decisions" (ENTSO-E, n.d.-a), as declared in its Mission statement. Thus, it has to be seen in part as a lobbying organization of the interests of TSOs and is registered in the official transparency register of the EU since February 2012 (European Commission, n.d.).

A.1.2. The European Union Agency for the Cooperation of Energy Regulators (ACER):

The ACER is the decentralized agency of the EU that was created for the Union-wide regulation and coordination of the European energy market.¹¹ It coordinates the cross-border cooperation among TSOs, DSOs and national regulatory agencies (NRAs). Thereby, it aims to mitigate the serious problems for the cross-border energy market that national regulatory measures are seen to cause if they are not properly coordinated internationally.¹² Other tasks are the counseling of other institutions and bodies of the EU on questions pertaining the energy system as well as the issuing of opinions and recommendations to other stakeholders, such as TSOs, DSOs and NRAs. Furthermore, the ACER has a supervisory role over other institutions such as the ENTSO-E, ensuring EU regulations are met. It has to work for the good of the EU, independently of private and business interests.¹³

¹¹Regulation (EC) No 714/2009 of the European Parliament and of the Council of 13 July 2009 on conditions for access to the network for cross-border exchanges in electricity and repealing Regulation (EC) No 1228/2003, 2009.

¹²Regulation (EU) 2019/942 of the European Parliament and of the Council of 5 June 2019 establishing a European Union Agency for the Cooperation of Energy Regulators (recast), 2019, Recitals 3-5.

¹³Regulation (EU) 2019/942 of the European Parliament and of the Council of 5 June 2019 establishing a European Union Agency for the Cooperation of Energy Regulators (recast), 2019, Article 1(3).

A.2. The legal foundations of the European Resource Adequacy Assessment

The legal framework of the ERAA is mostly defined in Chapter IV of Regulation (EU) 2019/943 of the European Parliament and of the Council of 5 June 2019 on the internal market for electricity (recast) (2019), in the following referred to as "Electricity Regulation". Article 23 of said regulation defines that the ERAA shall identify concerns regarding the adequacy of the electricity system to supply current and projected demands within the EU, on member state and bidding zone levels, over a 10-year period. The ENTSO-E is responsible for the annual conduction of the assessment. It is required to submit a draft of a methodology to the ACER by January 2020. The methodology should be transparent and consider various scenarios anticipating future developments while using probabilistic calculations and market models to ensure thorough and reliable risk estimations. The methodology and the results of the ERAAs shall be subject to a prior consultation of the member states of the EU, ACER and other relevant stakeholders. Article 27 of the Electricity Regulation determines that ACER has to decide if it can approve the methodology and ERAAs as they are, with amendments or not at all.

Furthermore, the regulation declares the ERAA to be the official basis for the evaluation of resource adequacy concerns in the EU. It may be complemented by national assessments that include additional sensitivities but follow the same general methodology. If resource adequacy concerns are identified, the affected states have to devise a plan for their elimination, that has to be approved by the Commission of the EU and monitored by annual reports.¹⁴ Only if these plans do not suffice to dismiss concerns, may so-called capacity mechanisms be introduced and only for a limited time of no more than 10 years.¹⁵ Capacity mechanisms are measures taken to ensure the availability of sufficient generation capacities to meet demand at all times, even if these capacities are not economically competitive. This is mostly done by subsidies to providers of generation capacities (Simoglou & Biskas, 2023).

¹⁴Regulation (EU) 2019/943 of the European Parliament and of the Council of 5 June 2019 on the internal market for electricity (recast), 2019, Article 20.

¹⁵Regulation (EU) 2019/943 of the European Parliament and of the Council of 5 June 2019 on the internal market for electricity (recast), 2019, Article 21.

A.3. Methodology

In October 2020 ACER approved an amended methodology for the European resource adequacy assessment, which serves as the foundation of the conduction and evaluation of the ERAAs (ACER, 2020). The assessment is centered around central reference scenarios that define projected demand, supply and grid assumptions collected by the TSOs. These scenarios are to be consistent with the NECPs, defining the national road maps for the energy transition and goals for greenhouse gas emissions. They should, furthermore, anticipate the impact of other policy measures concerning the energy system, particularly the plans for the elimination of resource adequacy concerns. One central reference scenario is then to consider capacity mechanisms, while another is to neglect them except for already awarded contracts (ACER, 2020, Article 4).

Given this input data, an EVA is to be conducted. The EVA aims to predict at what times capacity resources are removed, mothballed, reintroduced or added to the market based on their economic feasibility. This information is used to adjust the input data to have a more realistic estimation of the available technologies at a point in time (ACER, 2020, Article 6). The modified scenarios are used for an economic dispatch (ED) model, that determines what capacity resources are to be used to what extent for every hour of a target year. The ED model assumes perfect foresight of availability and demand time series and determines the dispatch of generation, storage, and demand units to meet demand while minimising the total operating costs of the system. For this, it has to account for various inherent constraints such as cross-zonal transmission capacities. It produces a variety of outputs notably the expected energy not served (EENS) and the loss of load expectation (LOLE) before and after activation of out-of-market capacities (ACER, 2020, Article 7).

Declaration of Authorship

English: I hereby declare in lieu of an oath that I have written this thesis independently and have not used any sources or aids than those specified. I also declare that I have followed the general principles of scientific work and publication as laid down in the guidelines of good scientific practice of the Carl von Ossietzky University of Oldenburg.

German: Hiermit versichere ich an Eides statt, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Außerdem versichere ich, dass ich die allgemeinen Prinzipien wissenschaftlicher Arbeit und Veröffentlichung, wie sie in den Leitlinien guter wissenschaftlicher Praxis der Carl von Ossietzky Universität Oldenburg festgelegt sind, befolgt habe.



Oldenburg, 20th June, 2025