UiT The Arctic University of Norway

Faculty of Science and Technology
Department of Physics and Technology

Enhancing Decision-making in the Electric Power Sector with Machine Learning and Optimization

Odin Foldvik Eikeland

A dissertation for the degree of Philosophiae Doctor – July 2023

Abstract

The electric power system infrastructure is essential for modern economies and societies, as it provides the electricity needed to power homes, businesses, and industries. It is of critical importance that the operation of the power system is optimized to serve the electricity demand reliably and sustainably. Advances in machine learning and optimization have enabled the potential to enhance decision-making in the electric power sector by gaining deep insight into the vast amount of data stored digitally.

The operation of electric power systems poses many challenges, such as those related to the rising integration of renewable energy sources, the need for energy storage, and the aging transmission infrastructure. To address these challenges, this thesis explores machine learning and optimization techniques to enhance decision-making concerning decarbonization targets, integration of renewable energy sources, cost savings, and reliable power supply.

The main contributions are the following:

- In the first work we presented a framework for predicting the electricity demand. We compared the accuracy of statistical and machine learning models at short- and medium-term forecasting horizons. The experimental results showed that machine learning methods achieve higher accuracy overall and exhibit good transferability, as they managed to predict the load at new locations that were not accounted for during training. The study highlights the importance of selecting the appropriate model to accurately predict the electricity demand in locations where historical consumption data may be limited or unavailable.

- Next, we analyzed the electricity transmission grid to identify the potential causes of disturbances in the power distribution network using machine learning classification techniques. Traditional classification methods can only indicate variables that, on average, mainly explain fault occurrences. In addition to providing such a global interpretation, it is essential to identify the specific variables that explain each fault. To address this challenge, we adopted a recent technique to interpret the

decision process of a deep learning model called Integrated Gradients. The proposed approach demonstrated the importance of gaining detailed insights into the occurrence of a specific fault.

- We adopted probabilistic forecasting to model the uncertainty when predicting electricity generation from wind power in the third work. As point forecasts do not account for such uncertainties, it is necessary to rely on probabilistic forecasts. We demonstrated how deep learning models can make day-ahead probabilistic forecasts and compared the accuracy of different dataset configurations. The findings show that the accuracy of forecasts improves when historical data on measured weather and numerical weather predictions are included as exogenous variables. This study shows the importance of understanding which covariates must be included in the dataset to improve the accuracy of the predictions.

- We modeled the electric power system using a novel optimization technique in the fourth and fifth works. In the fourth work, we analyzed the benefit of using a low-cost thermal energy storage unit called Thermal Energy Grid Storage (TEGS) to balance intermittent generation from solar energy systems. Our analysis emphasizes the need for storage to provide the grid with electricity more reliably. In the fifth and final work, we optimized the engineering design of TEGS to minimize the cost of decarbonization in electric power systems. The findings show that TEGS enables cost-effective grid decarbonization and improves reliability compared to a baseline scenario where TEGS is not an available technology.

List of publications

The thesis is based on the following original journal papers.

I O.F. Eikeland, F. M. Bianchi, H. Apostoleris, M. Hansen, Y. C. Chiou, and M. Chiesa. 2021. **Predicting Energy Demand in Semi-Remote Arctic Locations**", *Energies 14*, no. 4: 798, 2021, doi: https://doi.org/10.3390/en14040798.

II O. F. Eikeland, I. S. Holmstrand, S. Bakkejord, M. Chiesa, and F. M. Bianchi, "**Detecting and Interpreting Faults in Vulnerable Power Grids With Machine Learning**", *IEEE Access*, vol. 9, pp. 150686-150699, 2021, doi: 10.1109/ACCESS.2021.3127042.

III O. F. Eikeland, F. D. Hovem, T. E. Olsen, M. Chiesa, and F. M. Bianchi, "**Probabilistic forecasts of wind power generation in regions with complex topography using deep learning methods: An Arctic case**", *Energy Conversion and Management: X*, vol. 15, p. 100239, 2022, doi: https://doi.org/10.1016/j.ecmx.2022.100239.

IV O.F. Eikeland, C.C. Kelsall, K. Buznitsky, S. Verma, F. M. Bianchi, M. Chiesa, and A. Henry, " **Power availability of PV plus thermal batteries in real-world electric power grids**", *Applied Energy*, vol. 348, p. 121572, 2023, doi: https://doi.org/10.1016/j.apenergy.2023.121572.

V O.F. Eikeland, R. Macdonald, H. Apostoleris, S. Verma, K. Buznitsky, M. Chiesa, and A. Henry, "**Cost-Effective Thermal Energy Grid Storage for Decarbonizing Electric Power Systems**", Submitted to *Joule*, Jul. 2023.

Other papers
The following journal and conference papers also contribute to this thesis but are not included.

6. O.F. Eikeland, F. M. Bianchi, I. S. Holmstrand, S. Bakkejord, S. Santos, and M. Chiesa. 2022. "**Uncovering Contributing Factors to Interruptions in the Power Grid: An Arctic Case**", *Energies 15*, no. 1: 305. https://doi.org/10.3390/en15010305.

7. O. F. Eikeland, H. Apostoleris, S. Santos, K. Ingebrigtsen, T. Boström, and M. Chiesa, "**Rethinking the role of solar energy under location specific constraints**", *Energy, vol. 211*, p. 118838, 2020/11/15/ 2020, doi: https://doi.org/10.1016/j.energy.2020.118838.

8. O.F. Eikeland, F.M. Bianchi, I.S. Holmstrand, S. Bakkejord, M. Chiesa (2022). "**Detecting the Linear and Non-linear Causal Links for Disturbances in the Power Grid.**" In: Sanfilippo, F., Granmo, OC., Yayilgan, S.Y., Bajwa, I.S. (eds) Intelligent Technologies and Applications. INTAP 2021. Communications in Computer and Information Science, vol 1616. *Springer*, Cham. https://doi.org/10.1007/978-3-031-10525-8_26.

9. O. F. Eikeland, M. Chiesa, "**Renewable energy-based tourism for Asians in the Arctic: A perspective from Northern Norway**", *Routledge*, Asian Mobilities in a Changing Arctic, 73-84, 2021/10/20.

10. O. F. Eikeland, T. Boström and M. Chiesa, "**Investigation of Energy Yield on Tromsøya by Mapping Solar Potential in ArcGIS**", *2019 Nordic Workshop on Power and Industrial Electronics (NORPIE)*, 2019, pp. 1-8, doi: 10.1109/NORPIE55843.2019.9967829.

Acknowledgements

There are several people I would like to thank when finishing this thesis which marks the end of one journey and the beginning of another.

First, I would like to express my sincere gratitude to my supervisor, Matteo Chiesa, for his guidance, support, optimism, and patience during this journey. I am also grateful for Matteo's effort in connecting me with different collaborators that have made this thesis possible. I would also like to thank my co-supervisor, Filippo Maria Bianchi, for all the discussions and supervision on machine learning. Matteo and Filippo have deepened my understanding of the connection between machine learning and the electric power system.

I sincerely thank Dr. Asegun Henry for his supervision and for hosting me at his lab at MIT, and I thank the members of Dr. Henry's lab for all the useful discussions, including Kyle Buznitsky, Shomik Verma, Alina Lapotin, and Colin C. Kelsall. To Ruaridh Macdonald, thank you for your advice and helpful discussions on optimization and modeling of electric power systems. And thank you for all the time and effort you have spent on our joint research projects.

Thanks to all my collaborators on the Smart-Senja project for providing insight into the industry sector's challenges in the energy transition. I would especially thank Arva and Ishavskraft for the great collaboration. To Maritsa Kissamitaki, thank you for designing the Figures that are used in this thesis. Thanks to my other co-authors and the Renewable Energy group at UiT for interesting discussions and support. I would like to thank my committee members for taking the time to read my thesis and attend the defense.

I would also like to thank my family and friends for your support. I am so lucky to have a family that encourages my research, and I would never be here without you. Last but definitely not least, I would like to thank my dear wife Kayla for her constant support and encouragement. I am so lucky to have a wife who is genuinely interested in my work and likes to discuss it together.

Odin Foldvik Eikeland,
Tromsø, July 2023

Contents

List of Abbreviations

AI Artificial Intelligence

ANN Artificial Neural Network

CAES Compressed Air Storage

CEM Capacity Expansion Model

CNN Convolutional Neural Networks

CS Cruise Ship

DNN Deep Neural Network

DSO Distribution System Operator

ERCOT Electric Reliability Council of Texas

ERNN Elman Recurrent Neural Network

ESN Echo State Network

EU European Union

FNN Feedforward Neural Networks

GHG Greenhouse Gas

GRU Gated Recurrent Unit

IEA International Energy Agency

IG Integrated Gradients

IPCC The Intergovernmental Panel on Climate Change

ISO Independent System Operator

LinearSVC Linear Support Vector Classifier

LSTM Long Short-Term Memory

ML Machine learning

MLP Multi-Layer Perceptron

NN Neural Network

NSE Non-Served Energy

NWP Numerical Weather Predictions

PAF Power Availability Factor

PHS Pumped Hydroelectric Storage

PI Prediction Interval

RBF Radial Basis Function

RBFSVC Radial Basis Function Support Vector Classifier

RE Renewable Energy

ReLU Rectified Linear Unit

RNN Recurrent Neural Networks

SVM Support Vector Machine

TEGS Thermal Energy Grid Storage

TES Thermal Energy Storage

TPV Thermophotovoltaics

TSO Transmission System Operator

1 Introduction

1.1 Energy analytics

The global society is now in an era where we should transform the electric power infrastructure from being dependent on fossil-based resources such as oil, coal, and gas to be dependent on technologies with zero Greenhouse Gas (GHG) emissions associated with it (IRENA, 2022). The power sector's full transformation can potentially reduce the global GHG emissions by 25% (IPCC, 2014; EPA, 2011). The main motivation for reducing GHG emissions is global climate change, and the increased environmental challenges recently underpinned in the last Climate Change Assessment Report by The Intergovernmental Panel on Climate Change (IPCC) (Masson-Delmotte et al., 2021). This report claims that there is now "code red" for humanity due to the increased temperature and more extreme weather. Therefore, the report states the urgent need to reduce greenhouse gas emissions and transform society into a zero-emission one where we no longer depend on fossil fuels. This transition is highly complex and difficult for a global society dependent on fossil-based technologies, and multiple challenges must be addressed. Additionally, the transition should take place within 30 years, a short period for changing the infrastructure of a global society.

A vast amount of ongoing research and development (R& D) is enabling pathways and scenarios for how to make the energy transition toward decarbonization possible. The International Energy Agency (IEA) recently presented a roadmap for the global energy sector called "Net Zero by 2050" (Bouckaert et

al., 2021) explaining how transitioning to a net-zero energy system by 2050 could be possible. The European Commission has recently presented the "the European Green Deal". The Green Deal presents the roadmap for how the European Union (EU) should strive to be the first climate-neutral continent with no emissions of greenhouse gases by 2050 (EU, 2020). The Green Deal also aims to inspire the rest of the global society to perform such a green transformation.

Data analytics within the electric power system sector has gained increased attention due to the combination of advancements in data science and a rapidly increasing amount of data being generated in digital format. In-depth analytics of the vast amount of data continuously being recorded within the electric power system enables enhanced decision-making[1] processes to reach decarbonization targets and ensure a reliable power supply. Other examples of disciplines where data analytics has led to new insights and solutions to existing challenges include e.g., healthcare, finance, education, media, and manufacturing (Mikalsen, 2019; Dixon, Halperin, & Bilokon, 2020; Shah, Patel, Adesara, Hingu, & Shah, 2021; Athmaja, Hanumanthappa, & Kavitha, 2017; Wuest, Weimer, Irgens, & Thoben, 2016).

Data within the electric power sector has been logged over several decades; thus, a vast amount of data is recorded. However, the information in this data has been utilized historically to a lesser degree due to a lack of analytics knowledge (Karschnia, 2022). Gaining in-depth knowledge of the energy data being recorded using data analytics frameworks has shown great potential to visualize the problems that must be addressed in achieving global decarbonization goals. Several companies [2], and research institutions [3] consistently work on collecting and analyzing data within the electric power sector to provide insights about what steps must be taken to achieve decarbonization goals.

This research direction, in which machine learning and optimization techniques play a key role, is the main focus of this thesis and will be referred to as *energy analytics* hereafter. Energy analytics is a broad field involving data and analytical techniques to gain insights and make decisions about electricity generation, distribution, and consumption. This can include analyzing data

1. The term decision-making has been widely analyzed and discussed for decades (March & Olsen, 1976). In this thesis, we use a rational decision-making approach. Rational decisions involve identifying problems, gathering data, analyzing results to gain knowledge, and deciding what measures should be taken to reach actual aims.
2. Some examples include: DNV (https://www.dnv.com/), EIA (https://www.eia.gov/), IEA (https://www.iea.org/), IRENA (https://www.irena.org/), Rystad Energy (https://www.rystadenergy.com/).
3. Some research institutions focusing on the energy transition: MIT energy initiative (https://energy.mit.edu/), Arctic Centre for Sustainable Energy (ARC) (https://uit.no/research/arc).

from smart meters, sensor networks, and other sources to optimize electricity generation and usage and identifying patterns and trends that can help reduce costs and improve efficiency. Energy analytics can also identify potential issues, such as equipment failure, and develop strategies to mitigate risks and improve overall performance (Belagoune, Bali, Bakdi, Baadji, & Atif, 2021). Fig. 1.1 illustrates how energy analytics processes might look in practice.

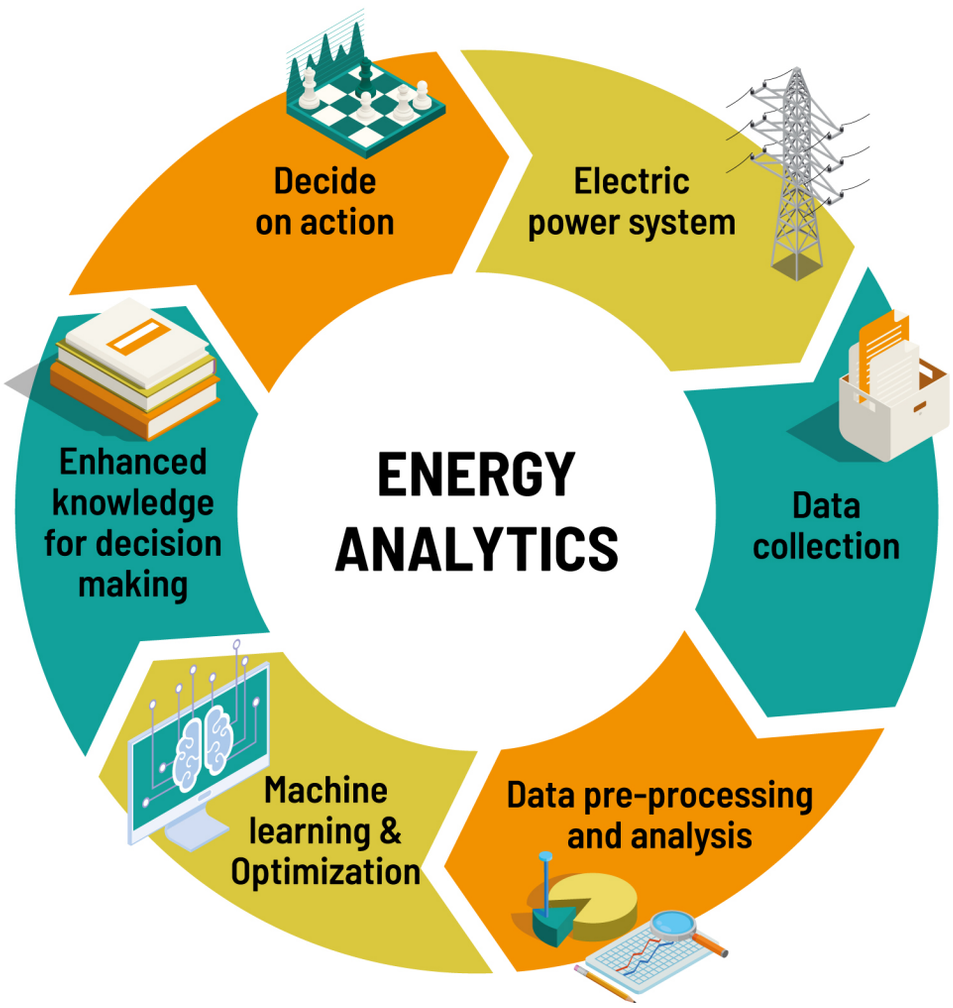


Figure 1.1: Overview of how Energy analytics might look like in practice. First, the data is collected from different parts of the electric power system. Then the collected data are pre-processed and analyzed before machine learning and optimization techniques are applied to gain insight into the collected data. The data insight gives enhanced knowledge to decide on the most optimum action to solve the problem.

The energy analytics research field is developing rapidly with respect to the increased attention to urgent decarbonization needs, as reflected by the vast number of companies and research institutions that work on these topics. As a result of these efforts, many research articles have been published in academic journals and reports, showing great promise for energy analytics approaches to give the insight to enhance decision-making in the electric power sector (Manfren, Nastasi, Groppi, & Garcia, 2020; Singh, Bocca, Gomez, Dahlke, & Bazilian, 2019; Bocca & Ashraf, 2022; Breyer et al., 2022).

A popular research field within energy analytics includes using machine learning methods to analyze and predict the electricity demand (Yildiz, Bilbao, & Sproul, 2017; Bouktif, Fiaz, Ouni, & Serhani, 2018; Hong & Fan, 2016; Bedi & Toshniwal, 2019; Chou & Tran, 2018). Similarly, the growing share of renewable energy sources in the electric power system has developed the research field of making accurate predictions of electric generation from technologies such as wind and solar energy, which has an intermittent generation profile (N. Sharma, Sharma, Irwin, & Shenoy, 2011; Kumar & Kalavathi, 2018; Foley, Leahy, Marvuglia, & McKeogh, 2012; Hossain, Chakrabortty, Elsawah, Gray, & Ryan, 2021; Zhang, Wang, & Wang, 2014; Mashlakov, Kuronen, Lensu, Kaarna, & Honkapuro, 2021). Predicting the electricity demand and generation accurately at different time horizons is critically important for several reasons. A few examples include resource planning (better plan how to utilize the available resources in the grid), grid stability (forecast peak demand to allocate resources to meet the need), cost optimization (using the most electricity in low-cost periods), renewable integration (asses the grid capacity to accommodate variable renewable energy electricity generation), energy market operations (market participants can maximize revenue by optimize generation schedules), as well as investment and planning decisions.

To optimize the reliability and energy security of the electric power system, it is necessary to detect and predict unscheduled power disturbances in the distribution network, which has severe consequences for both customers and grid operators. Several academic studies use machine learning methods to detect and predict such events with promising results (Owerko, Gama, & Ribeiro, 2018; Perera, Nik, Chen, Scartezzini, & Hong, 2020; Panteli & Mancarella, 2015; Hoffmann, Michałowska, Andresen, & Torsæter, 2019; Chen, Hu, Zhang, Yu, & He, 2019). Additionally, specific research initiatives are working on developing methodologies to predict faults in the grid [4]. Accurately detecting faults in the electric distribution grid enables utilities to enhance reliability, improve safety, allocate the available resources efficiently, and make better planning

4. This research initiative focuses on improved vegetation management and more resilient electric grids through satellites and artificial intelligence (https://www.stormgeo.com/solutions/data-science/grideyes/)

and investment decisions.

Another popular energy analytics approach is to use Capacity Expansion Model (CEM) to model electricity systems. Such models allow the construction of a digital representation of a real-world electric grid (i.e., a digital twin) and gain insight into how the electric power system should be operated in different scenarios. Such modeling exercises have become increasingly popular to gain insight into how to decarbonize electric power systems and how the portfolio of generation technologies in decarbonized grids should look like, as well as provide knowledge regarding what steps (i.e., cost improvement, technical specifics) that must be taken to achieve the modeled decarbonization goals (Sepulveda, Jenkins, Edington, Mallapragada, & Lester, 2021; Cole et al., 2021; Baik et al., 2021; Denholm et al., 2021; Sepulveda, Jenkins, de Sisternes, & Lester, 2018).

Nevertheless, despite the many promising results reported in academic journals and company reports, solutions implemented in industries [5], the seemingly large availability of data, the vast amounts of companies [6] and initiatives, and the great promises reported in media (Åge Algerøy, 2023; Dimmen, 2022; Bloomberg, 2023; Mehlum, Hischier, & Caine, 2021; Enel, 2023), data analytics using machine learning and optimization based-approaches within the power sector still have challenges that must be further addressed (i.e., data quality, analytics knowledge, data integration, scalability, data privacy and security). Thus, although energy analytics has been an active research field for many years, it is still research areas that should be further explored to enable insights to enhance decision-making within the power sector. In the next section, we will briefly describe some of the challenges and objectives of energy analytics approaches that will be addressed in this thesis.

1.2 Challenges and objectives of energy analytics

Despite the challenges and objectives of energy analytics being well-documented in the literature, the complexity of the electricity system, in general, makes it difficult to provide a complete overview of all challenges and

5. This industry harnessed Artificial Intelligence (AI) decarbonize their operations (https://www.nature.com/articles/d42473-021-00508-6). The Norwegian grid operation company Elvia is implementing Machine learning (ML) strategies to detect faults in their power grids (https://computas.com/elvia-og-computas-loser-samfunnsutfordringer-med-maskinlaering/)
6. e.g. some startup-companies focusing on applying AI to tackle climate change (https://www.forbes.com/sites/robtoews/2021/06/20/these-are-the-startups-applying-ai-to-tackle-climate-change/?sh=430b1647b26c)

objectives in the energy analytics field. In Fig. 1.2, we have provided a rough overview of some major elements.

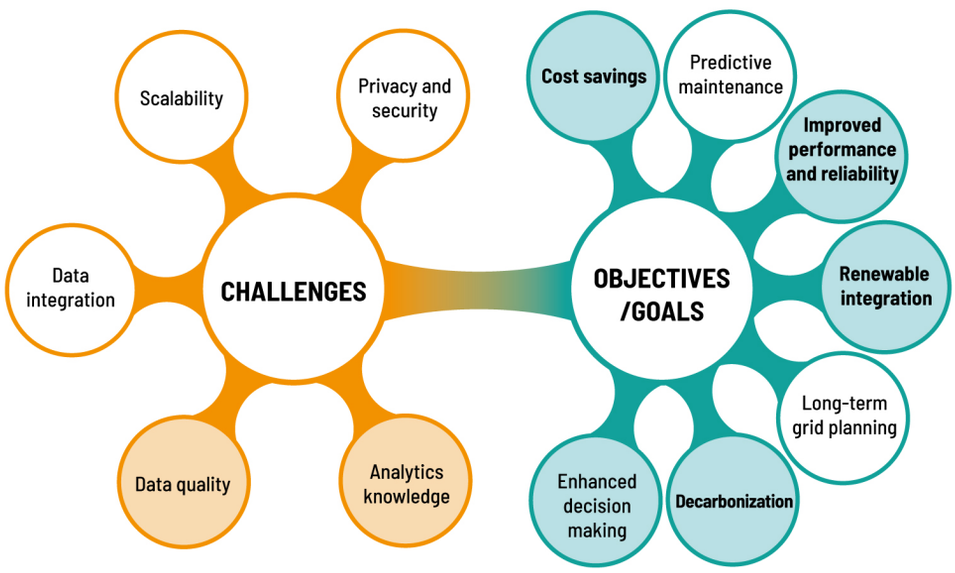


Figure 1.2: Overview of some challenges and objectives the energy analytics field faces. The elements that we are tackling in this thesis are marked in orange (challenges), and blue (objectives/goals).

Challenges

Data quality. Consistent and accurate data is critical to ensure precise results in an energy analytics process. It is, therefore, essential to gain a strong understanding of which data to include in the analysis and ensure that it contributes to insight into the problem. Hence, a good understanding of the problem at hand is important to collect data of sufficient quality that can explain the analysis results.

Data Integration. Energy systems are often highly distributed, with data coming from various sources, such as building management systems, smart meters, and sensor networks. Integrating and harmonizing data from these different sources can be difficult (Lopes, Hatziargyriou, Mutale, Djapic, & Jenkins, 2007; Almas, Vanfretti, Løvlund, & Gjerde, 2014).

Scalability. As the amount of data generated by energy systems continues to grow, the ability to process and analyze that data promptly and efficiently becomes increasingly important. Scalability is a major challenge for energy analytics processes (Bhattarai et al., 2019).

Privacy and Security. With the increasing use of smart meters and other Internet

of Things devices, there are concerns about the privacy and security of energy data. Ensuring the security of energy data and protecting consumers' privacy is an important challenge for energy analytics (Boroojeni, Amini, & Iyengar, 2017).

Analytics knowledge. Although a vast amount of data is continuously being logged within the energy sector, there is an urgent need to develop expertise in analyzing the available data to gain meaningful insights that can be used to enhance decision-making.

Goals and objectives
Cost Savings. Energy analytics can help organizations reduce costs by identifying inefficiencies and optimizing their operation.

Improved Performance and reliability. Energy analytics can be used to improve the performance of electric power systems, such as by identifying and addressing equipment failures, reducing downtime, and increasing reliability.

Renewable Integration. With the increasing penetration of renewable energy sources, energy analytics can be used to optimize the integration of renewable energy into the electric power grid and improve the grid's overall performance.

Predictive Maintenance. By analyzing data from electric power systems, energy analytics approaches can be used to predict when equipment is likely to fail, allowing for proactive maintenance and reducing downtime (Selcuk, 2017; De Benedetti, Leonardi, Messina, Santoro, & Vasilakos, 2018).

Decarbonization. Energy analytics can be used to monitor and identify energy consumption patterns and inefficiencies, which can help organizations reduce their carbon footprint and achieve decarbonization goals.

Enhanced decision-making. An in-depth analysis of the data sets is of value to gain insights into optimizing electricity usage, operation, and generation, which can help organizations make better-informed decisions about actions that must be taken.

1.3 Objectives and Achievements

In this thesis, we focus on some of the above-mentioned challenges and objectives. Our main achievement is providing methodological solutions that address the objective of *Improved performance and reliability of the power system.* All

five included papers deal with this objective.

The secondary achievements are to provide methodological solutions to objectives related with

- Renewable integration
- Decarbonization
- Cost savings

In addition, we also touch upon challenges related to ensuring high data quality by having an in-depth understanding of the task at hand and discuss why lack of analytics knowledge is an important challenge that must be addressed to enhance decision-making. Fig. 1.3 provides an overview of how the different publications relate to the objectives.

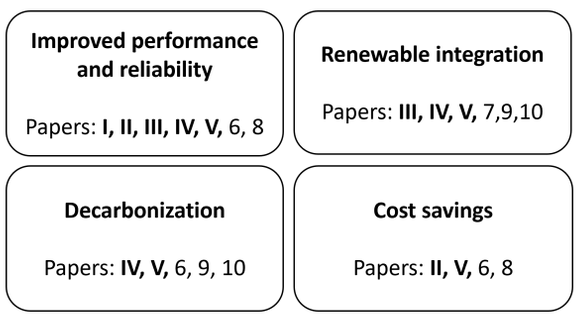


Figure 1.3: Categorization of publications according to the objectives they deal with.

By addressing these challenges and objectives using energy analytics approaches, this thesis aims to provide solutions to obtain the goal of *Enhanced decision-making* in the electric power sector.

1.4 Proposed approaches

The work presented in this thesis is motivated by the challenges the energy sector is facing; how to ensure a reliable electricity supply when transitioning the electricity system toward full decarbonization. The transformation from an electricity system mainly dependent on fossil-fuel-based technologies to an electricity system dominated by renewable energy technologies requires a deeper understanding of the data that is being logged in the energy sector. Advanced data science techniques such as machine learning and optimization have become promising approaches to gain additional insight into the data being logged, thus enable enhanced decision-making to decarbonize the electric

power sector. Using such techniques to analyze and gain deeper insight into the available data has also gained increased attention in other sectors, such as within the healthcare (Mikalsen, 2019) and finance sector (Dixon et al., 2020), to name a few examples.

In this thesis, we have analyzed datasets related to the energy sector. However, we take a general approach to these problems by using models that can be used in a wide range of applications. And thus, the approaches we come up with in this thesis are not restricted to applications within the energy sector only and can potentially be applied to any sector facing similar challenges. Our main goal in this thesis is to show the value of understanding the data available in a particular sector and use the increased knowledge to enhance decision-making.

In this thesis, the key solution to our main objective (*Improved performance and reliability of the power system*) is to use machine learning methods to make predictions and use optimization methods to identify cost-effective pathways to decarbonize the power system. In Paper **I** and Paper **III**, we used machine learning methods to predict electricity demand and generation. In Paper **I**, we used different machine learning models to predict the electricity demand, which is important when planning how to manage and optimize available energy resources. In Paper **III**, we analyzed the supply side of the electricity system by making probabilistic forecasts of the expected electricity generation from wind power plants. Wind power has a highly intermittent generation profile, so making probabilistic forecasts to account for uncertainties in a given prediction is important. Making accurate predictions that account for possible uncertainties is becoming important in the increasingly liberalized electricity market, where renewable energy contracts and auctions heavily rely on forecasting future power generation. This paper (Paper **III**) also addressed the challenge imposed by the increased share of *Renewable integration* in the electricity system.

It is of critical importance to have an electricity distribution grid that supplies consumers with electricity reliably. In Paper **II**, we addressed the problem of unscheduled power disturbances, which cause severe consequences for customers and grid operators. We used machine learning classification techniques to predict the onset of a fault. We adopted a recent technique to interpret the decision of the machine learning model, which allows for gaining detailed insights into the occurrence of a specific fault. Paper **II** also addresses the *cost savings* objective in the thesis, as identifying causes for failures in the power system can enable the distribution system operators to implement strategies to prevent and mitigate power disturbances which reduces the cost for both customers and the grid operators.

As variable renewable energy sources comprise a growing share of total electric-

ity generation, energy storage technologies are becoming increasingly critical for balancing energy generation and demand to ensure a reliable power supply to consumers. In Paper **IV** and Paper **V**, we address this challenge by using optimization techniques to model how storage can enable full decarbonization of emerging electric power systems that are heavily dependent on renewable energy technologies. Since the optimization schedule requires that we meet the demand for electricity at all times, Paper **IV** and Paper **V** also deal with the challenge related to *Improved performance and reliability*. A major challenge of the current energy storage technologies is the capital cost which is too high. Therefore, we model and optimize an existing thermal energy storage unit with estimated capital costs that are sufficiently low to enable large-scale deployment in the electric power system. Paper **IV** address the challenge of *decarbonization* and *renewable integration* since we model the electricity system under different scenarios where CO_2 emissions must be reduced. Such power systems are forced to depend more on renewable energy technologies such as wind and solar power. We deal with challenges *Cost savings*, *Renewable integration*, *Decarbonization*, and *Improved performance and reliability* in Paper **V**. Here we model hypothetical future decarbonized power systems that are dominated by wind and solar power technologies (*Decarbonization* and *Renewable integration*) and analyze how emerging storage technologies can be cost-optimized engineering-wise to reduce the cost for decarbonization (*Cost savings*).

1.5 Brief summary of papers

Paper I. In this paper, we present a framework for selecting the appropriate model by comparing the accuracy between statistical- and machine learning methods when predicting the electricity demand. We compare the prediction accuracy on several prediction horizons to gain insight into which model is most accurate on short-and medium-term load forecasts. Then we evaluated each model's transferability, which is important when predicting the demand where historical time series data may be limited. The findings show that the machine learning models achieved the most accurate predictions overall and exhibited good transferability as they managed to predict the load at different locations that were not accounted for during training.

Paper II. This paper presents a novel method to predict and interpret the causes of faults in the power distribution network. We compare linear and non-linear machine learning classification techniques to predict faults in an electric grid that experiences faults whose sources are unknown. We find that both the linear and non-linear classifiers achieved good classification performance, which shows that we constructed a data set that consists of features that explain well

the power disturbances. To gain detailed insight into which features explain specific fault occurrences, we adopted a recent technique to interpret the decision process of a deep learning model called Integrated Gradients. This is important knowledge for distribution system operators when implementing strategies to prevent and mitigate power disturbances.

Paper III. Improving prediction accuracy is of fundamental value for the energy market that relies on accurate forecasting capabilities. Here we addressed the problem of predicting the electricity generation from wind power that has a highly intermittent nature, increasing the uncertainty about future power generation. To account for the increased uncertainty, we computed probabilistic forecasts to generate samples of possible outcomes. Since the electricity generation from wind power heavily depends on external weather factors, we compared the prediction performance with different covariates. We found that the accuracy of the predictions improved vastly when we included the most optimum set of covariates. The work shows the importance of understanding which variables must be included to improve the prediction performance.

Paper IV. In this paper, we used a least-cost optimization model to analyze the value of using energy storage units to balance the electricity generation from solar power and thus provide reliable power to the grid. We modeled a thermal energy storage concept with estimated capital costs that are sufficiently low to enable large-scale deployment in the electric power system. The modeling was performed under a baseline case with no emission constraints and under hypothetical scenarios in which CO_2 emissions were reduced. The results show that the power available to the grid from our hypothetical solar and storage unit increases when the CO_2 emissions are reduced. The proposed approach shows how adding solar + storage systems to electric grids can contribute to the efficient stepwise decarbonization of electric power systems.

Paper V. In this paper, we build upon the work in Paper **IV**. We modeled the same storage concept based on thermal energy. In this work, we optimized the engineering design of the thermal energy storage to obtain the highest cost reduction compared with a baseline scenario where the storage unit is excluded. The findings show that the electric power system becomes approximately 3% cheaper than the baseline scenario for the thermal storage unit that is fully optimized engineering-wise. By modeling electric grids in different geographical regions, we find that the optimized engineering design depends on specific market conditions. The findings provide important insight into optimizing the value of emerging storage technologies to enable cost-effective decarbonization of the electric power system.

1.6 Organization of the thesis

The remainder of this thesis is organized into three parts; methodology and context, summary of research, and included papers. The methodology and context part contains two chapters. Chapter 2 describes the different sections of the electricity system and explains the value of in-depth analysis of the datasets generated in each power system sector. We also provide an example showing the value of using energy analytics approaches to identify ways to utilize renewables to cover a specific demand critical to decarbonizing the electric power system. Chapter 3 presents machine learning and optimization, which constitute the theoretical background for the research presented in this thesis. The machine learning time series section is relevant for Paper **I** and Paper **III**, machine learning classification is relevant for the work in Paper **II**, and the final section on optimization of electric power systems is relevant for Paper **IV** and Paper **V**.

In the summary of the research part, we provide a short overview of the scientific contribution of each paper in this thesis. We also add some concluding remarks, limitations of the works, and a discussion on future directions. The research papers are included in Part III of this thesis.

Part I

Methodology and context

/2

Energy Analytics: In-depth analysis of the electric power system

Electric power systems are complex infrastructures, including power generation and transmission grid lines that supply electricity to homes, industries, and other connected customers (Blume, 2016). The electric power system is a dynamic system that must be carefully managed to ensure a reliable supply of electricity to customers. This involves balancing the supply and demand of electricity in real-time, ensuring the system's stability, and maintaining a safe and secure power supply. Fig. 2.1 provides a schematic overview of the main elements of an electric power system.

In Fig. 2.1, electric power is generated using technologies such as Coal, Natural Gas, hydroelectricity, wind, and solar power. The generated electricity is then transported to customers through transmission grid lines, transporting high-voltage power over long distances. Power is transported to a substation that converts electricity to a lower voltage. Low-voltage power is directly transported to customers through the distribution grid lines. All elements in the electricity system must cooperate perfectly to avoid failures, resulting in critical power interruptions for customers. To reduce and mitigate the risk, it is critical to have an in-depth understanding of the data logged in different sections of the electricity system.

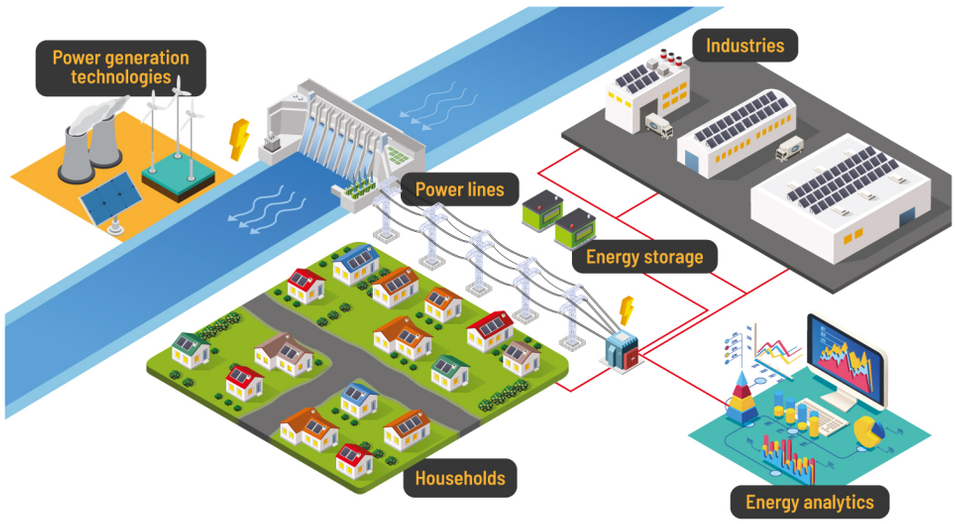


Figure 2.1: Illustration of different components of the electricity system. Electricity is generated using technologies such as e.g., solar, wind, natural gas, or hydropower before the electricity is transported through power lines. Then the power is further distributed to end users (e.g., households or industries). Energy storage technologies can optimize the operation of the electric power system by charging in periods with excess generation and discharging in periods with a surge in demand. Energy analytics is an essential part of the electricity system to gain insight into the generated data and thus enhance decision-making to optimize the operation.

To achieve global emission reduction targets, there is an urgent need to transform the electricity system from the current fossil-based system to an electricity system powered by Renewable Energy (RE) technologies with no associated emissions (IPCC, 2018). However, this transformation makes ensuring a reliable electricity supply to customers increasingly challenging owing to the increased dependency on RE technologies, such as wind and solar power, which have intermittent and uncontrollable generation profiles (Eikeland, Hovem, Olsen, Chiesa, & Bianchi, 2022; Zhou, Wang, Zhou, Clarke, & Edmonds, 2018; J. Yin, Molini, & Porporato, 2020). Consequently, to ensure reliable electricity supply in decarbonized power systems, there is an urgent need to understand better the operation of electric power systems that are increasingly dependent on RE technologies.

At the same time, as the electricity system has become increasingly complex to operate, there has been rapid development in the field of data science, where advanced machine learning and optimization techniques have become popular to gain more insight into energy systems (Mosavi et al., 2019; Donti & Kolter, 2021; Duchesne, Karangelos, & Wehenkel, 2020; Ahmad & Chen, 2020). Using such approaches to analyze energy-related datasets is referred to as Energy

Analytics which is the main topic of this thesis.

In the following, more detailed information on the different sections of the electricity system, with examples illustrating the importance of an in-depth analysis of the various sections. Such an analysis allows a better understanding of the available data within each sector. This is critical when deciding how to proceed by applying the correct machine learning and optimization methods to solve the task. Consequently, a good understanding of the available data enables the possibility to develop studies that can enhance the decision-making to optimize the operation of the electric power system.

2.1 Electric power generation

Electric power generation is the process of converting various forms of energy into electricity that can then be transmitted and distributed to customers (Grigsby, 2007). There are several ways to generate electricity, each with advantages and disadvantages. Some main types of power generation include fossil fuels (coal, natural gas, and oil), nuclear power, biomass, and renewable technologies, such as hydroelectric, biomass, wind, and solar power.

Reducing carbon emissions requires transforming from fossil fuel-based technologies to RE technologies with no associated emissions. The share of wind and solar energy in the electricity system has increased significantly and is believed to be the primary source of electricity supply in the future (Eriksen et al., 2022). In addition, to achieve zero emissions, the capital costs of wind and solar power have become lower than those of fossil-fuel-based technologies such as coal and natural gas, which has contributed to the rapid growth in installation in recent years (Kåberger, 2018; Apostoleris, Sgouridis, Stefancich, & Chiesa, 2018). In this thesis, we analyzed data from RE technologies that are believed to dominate future and decarbonized electric power systems. More specifically, we analyzed electricity generation data from wind and solar power to identify their applicability to play an increasing role in future decarbonized electricity systems.

The energy market relies on the demand and power generation forecasting capabilities that must be maintained in a dynamic balance. In Paper **III**, we tackled the challenge of predicting the expected electricity generation from wind power, which has a highly intermittent nature, using probabilistic forecasts. We performed a detailed analysis of the specific wind power plant we studied to understand better which features affect the generation of this particular power plant. In addition, we discussed with experts within the field and eventually constructed a dataset that consisted of the features that mainly contributed to

provide the most accurate prediction result.

Here we performed an in-depth analysis of the specific power plant and gained knowledge about which covariates must be included in the data set. Interestingly we achieved higher prediction performance once all the covariates were included. In addition, by including information regarding the technical limitations of the wind power plant, we gained even more accurate prediction results.

Another example where energy analytics were used to identify the applicability of using RE in an electric power system is described in Section 2.5. Here we tackled the challenge of intermittent electricity generation from solar energy by matching the generation profile to meet a specific demand for electricity.

2.2 Consumers of electricity

Electricity consumers are households, businesses, and organizations that use electricity to power their homes, buildings, and operations. Residential customers are individuals and households that use electricity to power their homes and appliances. They typically have a low electricity consumption compared to commercial and industrial customers, and the consumption follows a typical household behavior with high consumption during morning hours before the individuals are leaving for work, low consumption during mid-day, and an increased consumption during the afternoon due to cooking of dinner and family activities.

Industrial customers have a much greater electricity demand and have a different demand pattern than household consumers. Here energy consumption is affected by the need for the products and services the industry serves.

First, to gain better insight into the differences between the household and industry sectors, in Paper **I**, we transformed the datasets from the time domain to the frequency domain using Fourier transformation (Foldvik Eikeland et al., 2021a). In Paper **I**, we analyzed energy data from the household and industry sector to predict expected energy demand in both sectors, which is of essential importance for developing strategies to manage and optimize available energy resources. The uppermost graph in Fig. 2.2 shows the typical energy consumption frequency for the industry that was analyzed in Paper **I**. This industry has a peak frequency of 24 and 165 hours, which aligns with the industry's daily (24 hours) and weekly (165 hours) operations. On the other hand, the household frequency that is illustrated in the lowermost graph in Fig. 2.2 shows that the households that are analyzed have a primary frequency at the daily

cycle (24 hours), but also at 12 hours, which correlates well with the typical household everyday life with routines as breakfast every morning and dinner every afternoon. Analyzing the typical seasonality in electricity consumption using pre-processing techniques such as Fourier transformation is critical to understand the available data better and interpreting the results.

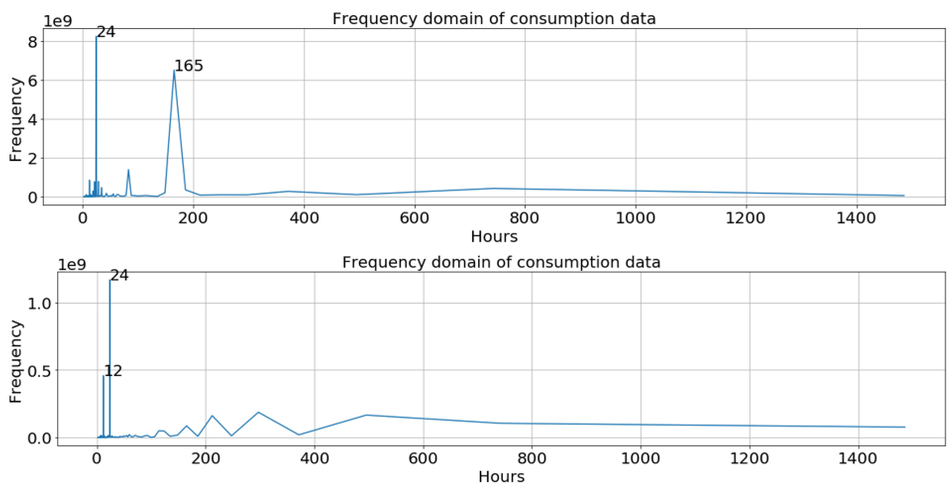


Figure 2.2: Difference in typical household and industry consumption periodicity. The uppermost graph indicates the industry sector, which clearly has a daily and weekly periodicity, while the lowermost figure shows the household sector has a clear daytime and daily periodicity. The frequencies for the industry and household sectors are from the time series in (Foldvik Eikeland et al., 2021b)

2.3 Energy storage

Energy storage technologies are units that store excess electricity to cover a demand at a later time. The goal of energy storage is to make energy available on request, regardless of energy availability from RE technologies. Therefore, to enable full decarbonization for the electric power system heavily dependent on RE sources, energy storage is an essential technology that must be a part of the energy transition to provide electricity to customers in a reliable way at all times. There are several types of energy storage technologies available, including Li-ion batteries, Pumped Hydroelectric Storage (PHS), Compressed Air Storage (CAES), Hydrogen, and Thermal Energy Storage (TES) (Armstrong et al., 2022).

However, a significant drawback for the existing energy storage technologies is that they are geographically constrained (PHS and CAES) or have a too

high capital cost which makes them unaffordable for the multi-day storage objectives required to completely decarbonize the grid (e.g., Li-ion batteries) (Braff, Mueller, & Trancik, 2016; Ziegler et al., 2019; Mallapragada, Sepulveda, & Jenkins, 2020). Studies suggest that achieving cost-efficient multi-day storage requires a capital cost reduction to US$3-30 kWh^{-1} (Ziegler et al., 2019; Albertus, Manser, & Litzelman, 2020).

This thesis addresses this problem by analyzing a storage concept based on Thermal Energy Storage TES, called Thermal Energy Grid Storage (TEGS). This technology has shown promising potential to achieve sufficiently low capital costs in the multi-day storage regime. TEGS stores the electricity as heat rather than electrochemically and then converts it back to electricity when needed (Henry, Prasher, & Majumdar, 2020).

To charge the TEGS unit, excess electricity is used to fuel resistive heating materials (graphite), transforming the electricity into heat at a temperature exceeding 2500°C. Then, the energy is transferred to graphite conduits via thermal radiation. Inside the conduits, liquid tin is used as the heat transfer fluid. The tin is heated from 1900°C to 2400°C, transforming the energy input into sensible heat and increasing its enthalpy. The liquid tin is continuously pumped through the conduits and then conveyed to the graphite blocks in the storage unit. When the 2400°C tin is pumped through the graphite blocks via conduits, it heats the graphite blocks from 1900°C to 2400°C via thermal radiation. Consequently, this cools the tin back to 1900°C. The tin is then reheated by being pumped back through the resistance heaters. This process constitutes the charging process until the graphite blocks are heated back to peak temperature. The storage unit should have a sufficiently large thermal mass to enable the storage unit to be charged for long periods with low heat loss (between 1% and 10%). The heat loss is an important design parameter which we optimize in Paper **V** with respect to minimize the cost of the TEGS unit. During discharging, liquid tin is pumped through the graphite storage to a power block. The power block consists of graphite conduits with unit cells. Each unit cell of piping creates a rectangular cavity lined with tungsten foil. This is a diffusion barrier to prevent graphite deposition onto Thermophotovoltaics (TPV) cells. Inside each cavity, the TPV cells can be lowered into the unit cell cavity. Here the TPV cells will be illuminated with the light emitted by the tungsten foil, which is heated by the light emitted by the graphite conduit carrying the tin. This net transfer of energy converts a large fraction ($> 50\%$) of the energy to electricity, which causes the tin's temperature to decrease to 1900°C before being pumped back to the graphite storage unit, where the tin is reheated again during the charging phase. In this way, the TEGS unit is a rechargeable grid-scale thermal battery that can store energy as heat and supply electricity to the grid on demand.

A recent development at lab-scale has enabled a roundtrip efficiency of 50% of the TEGS unit (LaPotin et al., 2022). The roundtrip efficiency of the TEGS unit is is entirely determined by the TPV conversion efficiency (i.e., the discharge efficiency from heat to electricity) (Amy, Seyf, Steiner, Friedman, & Henry, 2019; Kelsall, Buznitsky, & Henry, 2021). The charging efficiency (from heat to electricity) is assumed to be 100%. Technology developments of the TEGS unit at lab-scale has enabled the potential to achieve a projected cost below US$ 20 kWh^{-1} at gigawatt scales which is sufficiently low to completely decarbonize the electric power system (LaPotin et al., 2022; Amy et al., 2019; Kelsall et al., 2021). Fig. 2.3 provides a schematic illustration of the TEGS concept. A more detailed description of the TEGS technology is given in (Kelsall et al., 2021; Amy et al., 2019).

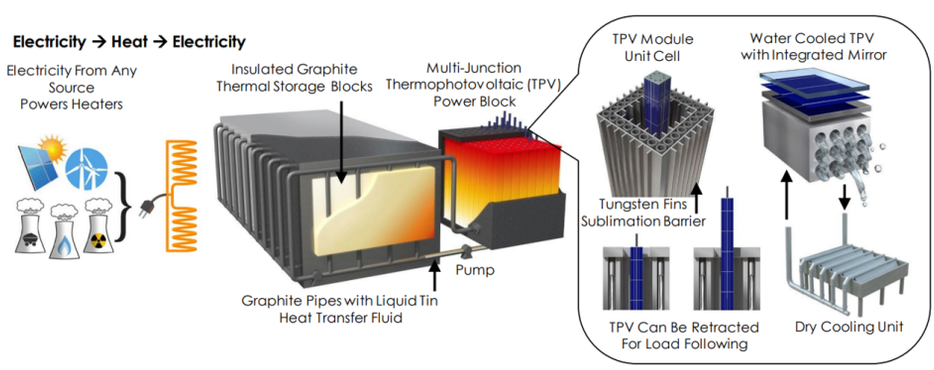


Figure 2.3: The TEGS technology concept. During charging, electricity from any generation technologies is used to power resistive heating materials (such as graphite or tungsten), transforming the electricity into heat at extremely high temperatures. During discharging, the energy (i.e., stored heat) is transferred to a power block consisting of Multi-Junction Thermophotovoltaic cells that convert the stored heat to electricity on demand. The illustration is from (Kelsall et al., 2021).

In Paper **IV** and Paper **V**, we studied how using TEGS can enable cost-efficient decarbonization of the electric power system. We gained an in-depth understanding of the techno-economic features of the TEGS concept. We combined the information about the capital cost and technical specifies with a CEM to investigate which engineering design is the most optimum one to utilize in the grid. We also find that for future electric power systems heavily dependent on weather conditions, the storage design requirements vary depending on the specific market conditions for each geographical region.

2.4 Transmission and distribution grid lines

The electric grid lines are divided into transmission and distribution lines. Transmission is high-voltage lines that transport electricity from power plants to sub-stations where the power is transformed to low-voltage lines, representing the distribution lines (Grigsby, 2007). The distribution system is the final step of electricity transportation towards the end-users. Grid lines play a crucial role in the electric power system. They allow electricity to be transported over long distances from power plants to end-users where the electricity is needed. They also connect different regions, sharing electricity across different states and regions. This makes the electric power system more flexible and resilient, allowing electricity to be moved from areas with surplus generation to areas with higher demand.

The Transmission and distribution lines are typically owned and operated by a Transmission System Operator (TSO), Distribution System Operator (DSO), or Independent System Operator (ISO) (Merino et al., 2021; Greer, 2012). They are responsible for the maintenance and operation of the grid. They also have a role in planning, developing, and expanding the grid lines. Transmission and distribution lines face challenges such as aging infrastructure and integrating more renewable energy sources. The transmission system needs to be upgraded and expanded to accommodate the growing demand for electricity (Clifford, 2023).

In this thesis, we focus on a specific grid in the Arctic region of North Norway, which faces problems with providing reliable power to its customers (Paper **II**, Paper 6, and Paper 8). More specifically, the customers are at the end of an old radial distribution network. The current distribution network has problems meeting the growing demand for electricity. This has increased the number of customer power interruptions, which has significant consequences for a community that relies on a stable power supply. A large portion of this thesis addresses the challenge of identifying potential causes for failure in the power system using energy analytics methodologies. In Paper 6, we collaborated closely with the DSO that operates the grid that was studied. We performed an in-depth grid analysis and were provided with inputs from the DSO, who had expertise and experience about potential causes for failures in the grid. Based on the research and information from experts, we constructed a dataset pertaining to the grid topology, the area's topography, the historical meteorological data, and the historical energy consumption/production data.

Gaining a deeper understanding of the problem to construct the correct dataset that could explain potential causes for grid failures proved an essential part of this study as we managed to predict the failure occurrence in the power grid with high performance. In Paper 8 and Paper **II**, we used a similar dataset

construction methodology as in Paper 6. The outcome showed that we also managed to predict fault occurrence using different prediction models.

2.5 Example of energy analytics applications

Lastly, in this section, we show an example of how energy analytics approaches can address a well-known problem of utilizing solar power in Arctic regions at high latitudes, namely the lack of solar availability during winter. However, there is a high solar availability during the summertime due to the midnight sun above the polar circle. Therefore, in Paper 7, we analyzed the yearly generation profile from solar energy to propose a system designed specifically for applications that are naturally matched to the solar resource. More specifically, we illustrated the value of using tailor-made photovoltaic systems to satisfy a specific Arctic demand: a booming cruise ship tourism industry in the north of Norway (Skinner, 2018). The Arctic has become an increasingly popular tourist destination in recent years, arguably due to 1) warmer temperatures and 2) increasing demand for the so-called "last chance" tourism as a consequence of the impact of climate change (Palma et al., 2019). Cruise Ship (CS) tourism is one of the fastest-growing economic sectors in some Arctic areas. The pollution associated with such growth of tourism activity aggravates the local air quality while increasing greenhouse gas emissions (D'Aprile, 2018).

To reduce emissions, the future CS industry is expected to retrofit ships to have the possibility to connect to shore power facilities. This could contribute to the electrification of CSs while harbored. In Paper 7, we proposed a system where CS is connected to shore power facilities that are powered by solar energy systems to provide renewable power to the CS that are harbored. Fig. 2.4 illustrate our proposed system design.

Our study concludes that solar energy generation could be a solid contribution to charging CS in the summer with no need for generation and transmission investments. Using energy analytics techniques enabled us to gain an in-depth understanding of the data concerning solar energy generation. We identified new ways of utilizing solar energy to cover a specific demand critical to decarbonize the electric power system.

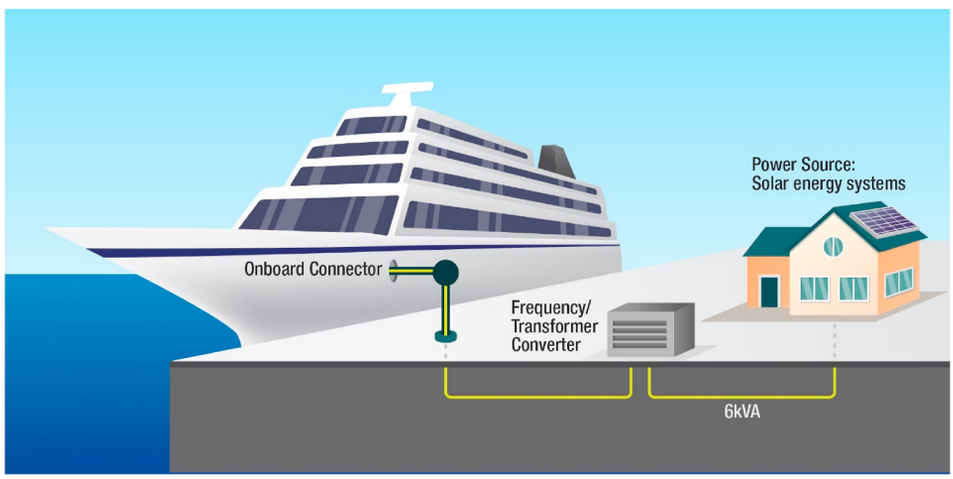


Figure 2.4: Typical system for connecting a ship to shore power when in port. This system consists of a power supply source where the energy is transported to a frequency/transformer converter to connect onboard. On the ship, the power is transformed and distributed to the different components from a control panel. The illustration is from our work in (Eikeland et al., 2020).

3

Gaining energy insight with machine learning and optimization

This chapter presents background theory on ML and optimization methods to gain insight into the energy-related datasets analyzed throughout the thesis papers. Machine learning is a subset of AI that allows models to train and learn from given input data and make predictions or decisions based on the specific task at hand (Alpaydin, 2020; Jordan & Mitchell, 2015). Several types of ML models have become popular to use in a wide range of applications. This thesis focuses on using different kinds of ML models to analyze energy-related datasets. Another data-science technique that has become popular to use to gain insight into energy-related data sets is optimization. Optimization techniques are methods to find the most optimum solution to a problem within a given set of constraints. Optimization techniques have become popular in various issues, including resource allocation, scheduling, transportation, and production planning. In this thesis, we are using least-cost optimization techniques to model the electric power system with respect to optimizing the resource allocation of electric generation technologies to achieve a power system with the lowest cost under different decarbonization scenarios.

3.1 Machine learning time series predictions

A time series is a collection of data points collected at regular intervals over time. Time series data is popular to analyze to identify patterns and trends in the data and to make predictions about future values based on past observations. This section presents background theory into ML time series predictions used in Paper **I** and Paper **II**. In this thesis, we have predicted the electricity demand from household and industry consumers and the electricity generation from wind power.

Several types of time series prediction methods exist, and two common techniques to predict time series are ML-based methods or statistical methods. Fig. 3.1 provides an overview of the different techniques used to predict the electricity demand in Paper **I**.

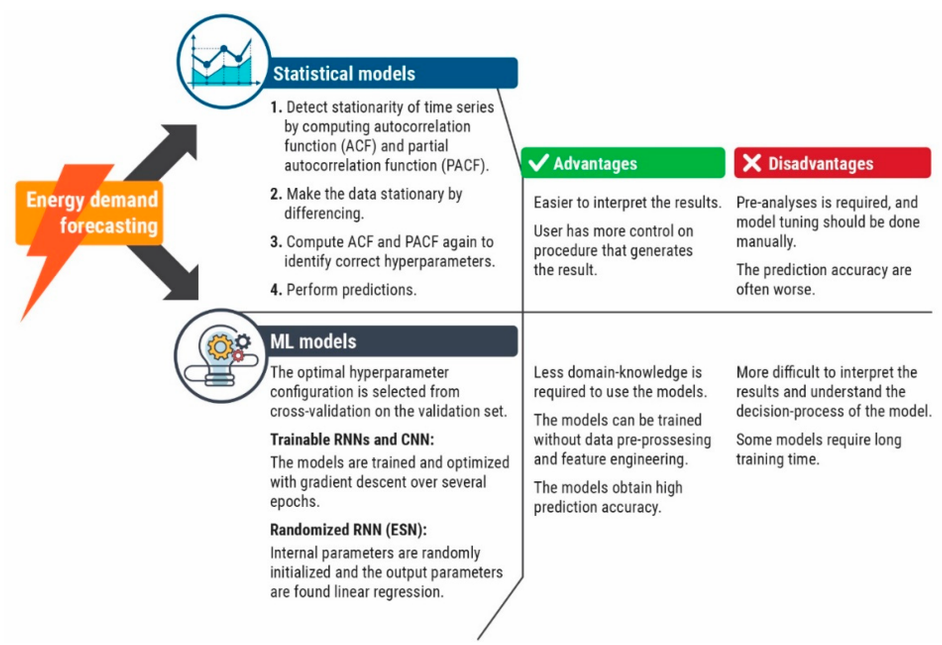


Figure 3.1: Overview of methodology for making time series predictions with statistical-and ML models. The advantages and disadvantages of each approach are given. The illustration is from our work in (Foldvik Eikeland et al., 2021b).

Our work in Paper **I** found that ML techniques outperform statistical methods when making time series predictions on energy demand data. Other work in the literature also highlights that ML models are better suited to make accurate predictions of time series of energy data (Gasparin, Lukovic, & Alippi, 2022; Bianchi, Maiorino, Kampffmeyer, Rizzi, & Jenssen, 2017). Therefore, this thesis

focuses on time series prediction using ML methods. More specifically, we use Artificial Neural Network (ANN), a ML architecture that has shown promise in making accurate predictions of time series data.

3.1.1 Artificial Neural Networks

ANNs are machine learning networks composed of layers of interconnected nodes, or neurons, that process and transmit information. The Neural Network (NN) is divided into three layers: The input, hidden, and output layers (see schematic overview in Fig. 3.2). The input layer provides the input data that the ANN should train on. In the hidden layer, the ANN learns from the given input data and trains the ANN algorithm by adjusting its connections and weights. Once trained, the network makes predictions or decisions based on the given task in the output layer. An ANN with several hidden layers is called a Deep Neural Network (DNN). Fig. 3.2 provides a schematic overview of a NN with three hidden layers.

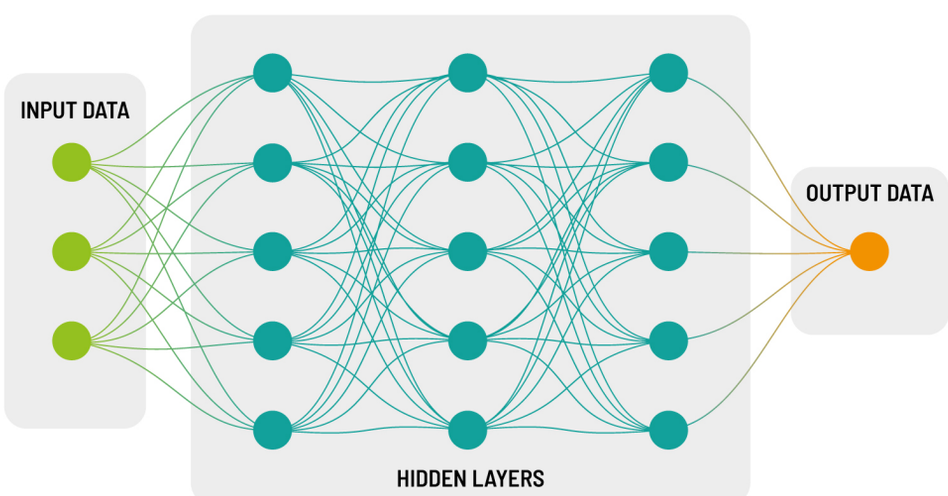


Figure 3.2: Schematic overview of a typical artificial neural network. First, input data is given. In the hidden layer, the NN processes the input data by applying weights to the inputs and produces outputs through an activation function (e.g., sigmoid, ReLU, or tanh). A NN can have one or more layers depending on the specific task. An NN with two or more hidden layers is referred to as a deep neural network. The output layer is where the result is presented for the particular problem.

ANNs are typically divided into Feedforward Neural Networks (FNN), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN). The papers included in this thesis mainly focus on using FNNs and RNNs to analyze energy

datasets. Therefore, we do not include a discussion about CNNs in this chapter. More information about CNNs can be found in (Albawi, Mohammed, & Al-Zawi, 2017), where the authors describe the CNN model. In addition, the authors in (Gu et al., 2018; Li, Liu, Yang, Peng, & Zhou, 2021) provided a comprehensive survey of recent development and applications of CNNs.

In a **feedforward NN**, one or more layers of neurons are connected to the next layer in a single direction without feedback connections. The input data are fed into the first layer, which processes it and passes it to the next layer (Bebis & Georgiopoulos, 1994). The output of the last layer is the final output of the network. FNNs are widely used for tasks such as classification and regression, in which the input is mapped to a corresponding output. The error between the predicted and actual output is calculated and used to adjust the weights of the connections between the neurons. The network is trained using a process called backpropagation (Sazli, 2006). The architecture of an FNN is defined by the number of layers, the number of neurons in each layer, and the activation function. The FNNs' most common activation functions are sigmoid, Rectified Linear Unit (ReLU), and hyperbolic tangent (tanh) (S. Sharma, Sharma, & Athaiya, 2017). The advantages of FNNs include their ability to learn complex nonlinear relationships between inputs and outputs, their ability to generalize new data, and their flexibility regarding the number of layers and neurons. However, they can be prone to overfitting if the network is too complex or the training data are too limited (Schittenkopf, Deco, & Brauer, 1997).

RNNs were designed to handle data sequences like time series, where each input has a temporal relationship with the previous inputs. Unlike FNNs, RNNs have feedback connections that allow them to pass information from one time step to another (Medsker & Jain, 2001). The key feature of an RNN is the presence of a hidden state that acts as a memory of the previous inputs. The hidden state is updated at each time step using a recurrent function that uses the current input and the previous hidden state as the input (Bianchi et al., 2017). The output of the RNNat each time step is a function of the current input and hidden states.

RNNs are helpful for tasks such as language modeling (Graves, 2013), speech recognition (Graves, 2011), and machine translation (Cho et al., 2014), where the output depends on the context of previous inputs. In addition, they have become increasingly popular for use in time-series prediction tasks because they provide state-of-the-art performance (Salinas, Flunkert, Gasthaus, & Januschowski, 2020; Bianchi et al., 2017). The effectiveness of RNNs in handling time series comes from their ability to learn an input sequence using a recurrent function. They can also be trained using backpropagation through time, a variation of the backpropagation algorithm used in FNNs (Rumelhart,

Hinton, & Williams, 1985).

In this thesis, the RNNs have achieved the best overall performance in our prediction tasks, and we hereby provide more detailed descriptions of a few popular models we have used in our thesis.

Elman Recurrent Neural Network (ERNN) is the most standard RNN (Gasparin et al., 2022). The ERNN was proposed by Jeffrey L. Elman (Elman, 1990). Here, one sequence element is processed simultaneously. At each timestamp t, the input layer processes information at $\mathbf{x}$[t] $\epsilon \mathbb{R}^{N_i}$, where N_i is the number of nodes in the input layer. The input time series $\mathbf{x}$ has a total length of T. In the input layer, each component is summed by a bias vector $\mathbf{b}_i \epsilon \mathbb{R}^{N_h}$, where N_h is the number of nodes in the hidden layer. Each component **x[t]** is then multiplied by the weight matrix $\mathbf{W}_i^h \epsilon \mathbb{R}^{N_i \times N_h}$. Similarly, the internal state $\mathbf{h}$[t-1] $\epsilon \mathbb{R}^{N_h}$ from the recurrent time is summarized using a bias vector $\mathbf{b_h} \epsilon \mathbb{R}^{N_h}$ before multiplying with the weight matrix $\mathbf{W}_h^h \epsilon \mathbb{R}^{N_i \times N_h}$ of the recurrent connections. The transformed input and past network state are then combined and processed by the neurons in the hidden layers (Foldvik Eikeland et al., 2021b). Finally, the output of the network at timestamp t is

$$\mathbf{y}[t] = g(\mathbf{W}_h^o(\mathbf{h}[t] + \mathbf{b}_o)), \tag{3.1}$$

where the output is computed through the transformation $g(\cdot)$ of the matrix of the output weights $\mathbf{W}_h^o \epsilon \mathbb{R}^{N_r \times N_o}$ (N_o is the number of nodes in the output layer). The output weights are applied to the the current state **h[t]** sum, and bias vector $\mathbf{b}_o \epsilon \mathbb{R}^{N_o}$ (Bianchi et al., 2017). One limitation of ERNNs is the vanishing gradient problem (Hu, Huber, Anumula, & Liu, 2018), in which the gradient of the loss function with respect to the weights becomes very small as the sequence length increases. This makes it difficult to train RNNs to capture long-term dependencies. Several variations of RNNs have been proposed to address this problem, including Long Short-Term Memory (LSTM) (Hochreiter & Schmidhuber, 1997) and gated recurrent units Gated Recurrent Unit (GRU) (Chung, Gulcehre, Cho, & Bengio, 2014), which use specialized recurrent units that can retain information over longer periods.

The LSTM architecture is comparable to the ERNN architecture. The composition of the inner module, where LSTM implements a more sophisticated internal processing unit called a cell, is the primary distinction between LSTM and ERNN (Bianchi et al., 2017; Gasparin et al., 2022). The LSTM uses a gated system to manage the information in this case. The ability of gated networks to resolve the vanishing gradient problem by not imposing any bias toward recent observations is its main characteristic and the reason they are so popular. As a result, the LSTM may keep its internal memory unchanged for extended

periods (Bianchi et al., 2017; Gasparin et al., 2022)

GRUs are simplified versions of LSTMs. In contrast to LSTM networks, GRU networks combine and merge the forget and input gates into a single update. This update controls the degree to which each hidden unit could remember or forget. In contrast to the LSTM, which has three gates, the GRU network only has two gates (Bianchi et al., 2017; Gasparin et al., 2022). Studies have demonstrated that GRUs can perform similarly to LSTMs but typically train more quickly due to less computational intensity (Chung et al., 2014; W. Yin, Kann, Yu, & Schütze, 2017).

Probabilistic time series forecasting

Making probabilistic forecasts to account for possible uncertainties in a given prediction has received increasing attention in the literature (Hong & Fan, 2016; Zhang et al., 2014; Mashlakov et al., 2021). In Paper **III**, we used a special architecture of RNNs to compute a probabilistic forecast of expected electricity generation from wind power. More specifically, we employed an autoregressive RNN called DeepAR proposed by (Salinas et al., 2020). The DeepAR model learns a global model from the historical data of all time series in the dataset and produces probabilistic forecasts by incorporating a Negative Binomial likelihood for count data. The internal units in the DeepAR model that process the input data can be LSTM or GRU. Fig. 3.3 provides a schematic overview of the DeepAR model with GRU or LSTM as internal units.

3.2 Machine learning classification

Instead of time-series predictions, it is possible to predict the class or category of an observation based on a set of input features. Such predictions are referred to as classification (Kotsiantis, Zaharakis, Pintelas, et al., 2007). This section presents the background theory of the ML classification techniques used in Papers **II**, 6, and 8. In these studies, we used ML classification models to predict the onset of faults in the distribution grid lines. In these studies, the classification datasets consisted of two labels: one representing a fault occurrence and one representing normal conditions where the grid operates as it should.

When classifying datasets, the classification can be linear or nonlinear. The differences between linear and nonlinear classification are shown in Fig. 3.4.

Linear classification works by finding a linear boundary (hyperplane) separat-

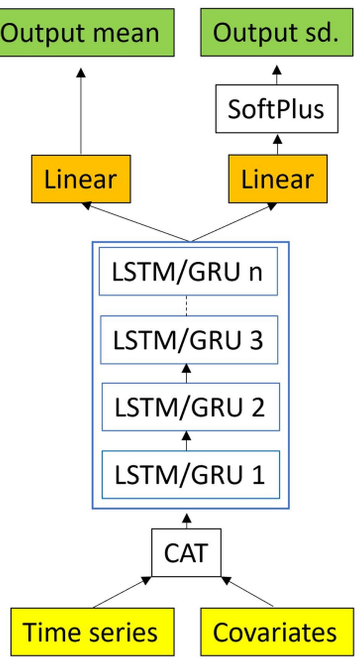


Figure 3.3: The DeepAR architecture. Here the time series and covariates are concatenated (CAT) and fed into the internal units (LSTM or GRU). The internal units can have one or more layers depending on which architecture gives the highest prediction performance. The output layer's mean and standard deviation are computed to generate predictive samples. The illustration is from (Eikeland, et al., 2022).

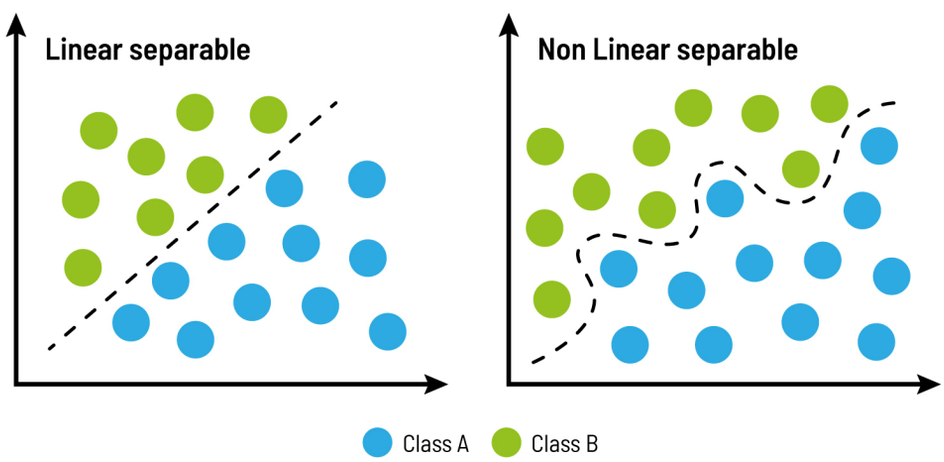


Figure 3.4: Difference between classes that are linearly and non-linearly separable.

ing different classes of input data (Konishi, 2014). In linear classification, the input data are represented as a set of features, and each feature is assigned a

weight. The model learns the optimal weights that maximize the separation between different classes. The linear hyperplane can then be used to predict the class of the new data points. Linear models are useful because they are used to construct a decision boundary directly into the input space. This allows the interpretation of the decision process of the classifier (Eikeland, Bianchi, Holmstrand, Bakkejord, & Chiesa, 2022). Nonlinear classification is when the hyperplane between classes is not linear. Nonlinear classification models are required when input data are complex and cannot be separated by a linear hyperplane. These models can capture more complex patterns and relationships in the data, allowing for more accurate classification if the input data are complex and cannot be separated linearly. However, such complex nonlinear relationships in the data can make the interpretation of the decision process of the classifier model more difficult (Montavon, Lapuschkin, Binder, Samek, & Müller, 2017). In this thesis, we adopt recent techniques to interpret the decision process of nonlinear classifiers.

Several types of linear and nonlinear classification techniques are popularly used in ML, and they are outside the scope of this thesis to describe all ML classification models. In this thesis, we provide background theory for the classification methods mainly used in Papers **II**, 6, and 8. The linear and non-linear classification models that were in this thesis are given in Tab. 3.1

Table 3.1: Classification models used in this thesis

Linear classifiers
Logistic Regression
Ridge Classifier
Linear Support Vector Classifier (LinearSVC)
Non-linear classifiers
Radial Basis Function SVC (RBFSVC)
Multi-level Perceptron (MLP)

The following sections provide background theory into the different models.

3.2.1 Linear classifiers

Logistic regression models use a logistic function to approximate the probability of a binary outcome (i.e., whether an event occurs or not). The Logistic regression model is used to classify data points into one or two distinct classes based on a set of input features (Bishop & Nasrabadi, 2006; Harrell et al., 2001). Logistic regression multiplies the input features by their respective weights or coefficients. The resulting linear combination is passed through a logistic (sig-

moid) function that maps the output to a probability value between 0 and 1. The logistic function was computed as follows:

$$P = \frac{1}{1 + e^{-x}}, \tag{3.2}$$

The maximum likelihood estimation function is a popular metric for computing the conditional probability for classification problems. The logistic functions show that if probability P exceeds 0.5, the predictions are classified as class 0. Otherwise, Class 1 is assigned (Belyadi & Haghighat, 2021).

Ridge Classifiers converts the target values into {-1, 1} and then treat the problem as a regression task (Bishop & Nasrabadi, 2006). Subsequently, a ridge regression model is constructed to predict the target variable. Unlike Logistic regression, where the maximum likelihood estimation is the loss function, the ridge model uses a mean squared error loss function with L2 regularization as a penalty term. The Ridge regression model is also known as the Tikhonov regularization (Golub, Hansen, & O'Leary, 1999). If the predicted value is greater than zero, the resulting class is +1, and the negative predicted class is assigned as -1.

Linear Support Vector Classifier (LinearSVC) is a special type of a Support Vector Machine (SVM) endowed with a linear kernel (Boser, Guyon, & Vapnik, 1992). the SVM aims to find a hyperplane in an N-dimensional space that distinguishes data points (Suthaharan, 2016). When separating classes, many possible hyperplanes can be selected. The objective of SVM is to find the hyperplane with the maximum distance between the data points of both classes. The SVM loss function that is computed to maximize the distance is computed using hinge loss as follows:

$$l = max(0, 1 - y^i(x^i - b)), \tag{3.3}$$

where y^i and x^i refer to the ith instance in the training set and **b** refers to the bias term. The objective of loss function l is to maximize the loss function, which maximizes the distance between classes. The LinearSVC model finds a linear hyperplane between two classes.

3.2.2 Non-linear classifiers

Radial Basis Function Support Vector Classifier (RBFSVC) is a special type of a SVM endowed with a non-linear Radial Basis Function (RBF) kernel. In contrast to SVM with a linear kernel, RBF can compute nonlinear relationships between two classes (Orr et al., 1996). The RBF kernel is widely used due

to its similarity to the Gaussian distribution (Musavi, Ahmed, Chan, Faris, & Hummels, 1992). The RBF kernel function is computed as the distance between two points X_1 and X_2 and is mathematically defined as follows:

$$K(X_1, X_2) = exp(-\frac{|| X_1 - X_2 ||}{2\sigma^2}), \tag{3.4}$$

where σ is the variance, and $|| X_1 - X_2 ||$ represent the Eucledian Distance between the two points X_1 and X_2.

Multi-Layer Perceptron (MLP) is essentially a FNN that generates a set of outputs from a set of inputs (LeCun, Bengio, & Hinton, 2015), as explained in previous sections. In paper **II**, we employed an MLP specifically constructed for classification problems (Eikeland, Holmstrand, Bakkejord, Chiesa, & Bianchi, 2021). This MLP consists of an input layer that takes the input vectors $x \epsilon \mathbb{R}^n$, L hidden blocks, an output layer that generates a 2-dimensional output $o \epsilon \mathbb{R}^2$, and softmax activation that gives the vector of class probabilities y. The MLP was trained by minimizing the cross-entropy loss using batches of size b and the Adam optimization algorithm (Kingma & Ba, 2014) with an initial learning rate r. We refer to Figure 1 in our Paper **II** for a graphical illustration of the MLP architecture (Eikeland et al., 2021).

3.2.3 Interpreting the decision process of non-linear classification models

It is difficult to interpret the decision process of a NN due to the presence of many non-linear transformations. However, we are interested in identifying the specific variables that explain each fault when predicting faults in the distribution grid lines. Thus, we can better understand which measures must be taken to prevent and mitigate specific power disturbances.

A considerable research effort has been devoted to understanding what a NN learns and thus makes decisions. A popular interpretation technique that aims to identify which inputs influence the model the most, is Gradient-based approaches. In (Simonyan, Vedaldi, & Zisserman, 2013), a saliency map was computed by taking the gradient of the class activation score with respect to each input feature. To project the activations of an intermediate hidden layer back to the input space, the authors in (Zeiler & Fergus, 2014) inverted a CNN operation from the hidden layer to the input layer. This allows gaining an insight into which details the hidden layer has captured from the input. Another gradient-based technique was proposed by authors in (Springenberg, Dosovitskiy, Brox, & Riedmiller, 2014). The authors constructed a Guided Back Propagation technique, outperforming the standard gradient backpropagation. However, when ReLU is encountered, the gradient is back propagated only if both the gradient

and the ReLU activation function are positive.

A drawback of gradient-based methods is that such approaches attribute zero contribution to inputs that saturate the ReLU or MaxPool activation functions. The authors in (Bach et al., 2015) introduced an axiom to capture such shortcomings of the gradient-based methods. The axiom state that the sum of the relevance of all pixels must be equal to the class score of the model, and the total relevance of the class score to the input features are distributed with a method called Layer-wise Relevance Propagation. A downside of this method is that it does not formalize how to distribute the relevance among the input features. The DeepLiFT method proposed by the authors in (Shrikumar, Greenside, & Kundaje, 2017) enforced an additional axiom on propagating the relevance by following the chain rule like gradients.

In Paper **II**, we use an Integrated Gradients (IG) technique proposed by (Sundararajan, Taly, & Yan, 2017). The IG is constructed to satisfy two axioms; sensitivity and invariance. Although the Layer-wise Relevance Propagation and the Deeplift approaches ensure sensitivity, their coarse approximations to gradients might break the invariance assumption.

In the following, we illustrate the working mechanism of the IG method by studying pixels (image classification). Here, we use the IG method to detect polar lows in a satellite image. We illustrate the IG mechanism using pixels as it is easier to visualize how the interpretation method works. In our Paper **II**, we used IG to interpret the decision process of an NN used to classify time series data concerning electric grid lines.

The IG method is divided into four steps:

1. Start from an uninformative **baseline** (e.g., a black image)

2. Interpolate small steps along a straight line in the feature space between the baseline and the actual image

3. Compute gradients at each step between the model's predictions with respect to each step

4. Approximate the integral between baseline and input by accumulating the local gradients

Step 1 and step 2 are illustrated in Fig. 3.5, where one starts from the black baseline image to the right and performs a linear interpolation to the actual image.

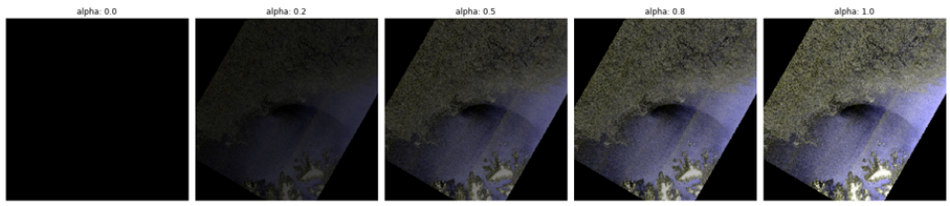


Figure 3.5: Interpolate small steps along a straight line in the feature space between the baseline and the actual image.

In step 3, the IG for each feature i is computed as

$$\text{IG}_i(x) ::= (x_i - x_i') \times \int_{\alpha=0}^{1} \frac{\partial F\left(x' + \alpha \times (x - x')\right)}{\partial x_i} d\alpha, \tag{3.5}$$

here i is an input feature, x is a sample in the dataset, x' is the uninformative baseline, and α is an interpolation constant used to perturb the features of the input sample. The above definition ensures the two axiom assumptions:

- Sensitivity: IG sum up the difference in feature score
- Invariance: IG attribution is completely defined in terms of gradients

As the computation of the integral is often not tractable, it is necessary to use numerical approximations. In step 4, the IG is approximated using the Riemann trapezoidal, which modifies the original IG integral into:

$$\text{IG}_i^{\text{approx}}(x) ::= (x_i - x_i') \times \sum_{k=1}^{m} \frac{\partial F\left(x' + \frac{k}{m} \times (x - x')\right)}{\partial x_i} \times \frac{1}{m}, \tag{3.6}$$

where m is the number of finite steps that approximate the integral and $\alpha \approx k/m$. The m samples $\mathcal{X} = \{x' + \frac{k}{m} \times (x - x')\}_{k=1}^{m}$ represent the linear interpolation between the baseline and the input (Eikeland et al., 2021).

Fig. 3.6 illustrate how IG is used to identify polar lows in satellite images (highlighted in green).

Due to its wide application to different NNs, simplicity of use, theoretical explanations, and computing efficiency, IG has become a popular interpretability method (Eikeland et al., 2021; Qi, Khorram, & Li, 2019; Sundararajan et al., 2017). In Paper **II**, we used the IG technique to detect potential causes for specific failures in the grid. Instead of using IG to analyze images, we analyzed

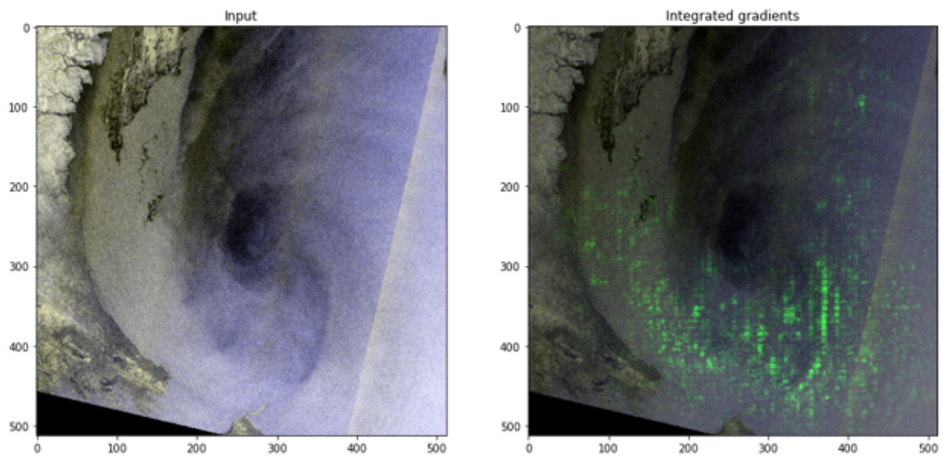


Figure 3.6: IG to identify polar lows in satellite images. The polar lows are highlighted in green color.

feature vectors that could potentially explain fault occurrence in the electricity distribution grid. A black image (all pixels at 0, as in the leftmost picture in Fig. 3.5) is commonly used as a baseline in computer vision tasks. However, when identifying potential causes for a specific fault, the value 0 might be informative because the absence of specific features can increase the probability of belonging to a specific class (e.g., in the absence of wind, it is less likely to observe a fault). From our Paper **II**, we found that the best method to find the baseline is to use a mean baseline, which gives almost the same probability of obtaining classes 0 or 1. The baseline x'_m is a vector computed as a weighted average of the features across the two classes. Fig. 3.7 depicts the interpolation path from the mean baseline to a specific class "fault" sample in the dataset.

3.2.4 Confusion matrix and the F1 score

The confusion matrix is a widely used metric for measuring the performance of classification models (Pedregosa et al., 2011). The confusion matrix summarizes the performance of the classification algorithm by comparing the predicted class labels with true class labels. The confusion matrix has four quadrants: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). The following is a common format for a confusion matrix:

True Positive (TP): The number of correctly classified as positive observations.
False Positive (FP): The number of incorrect observations classified as positive.
True Negative (TN): The number of correctly classified as negative observations.

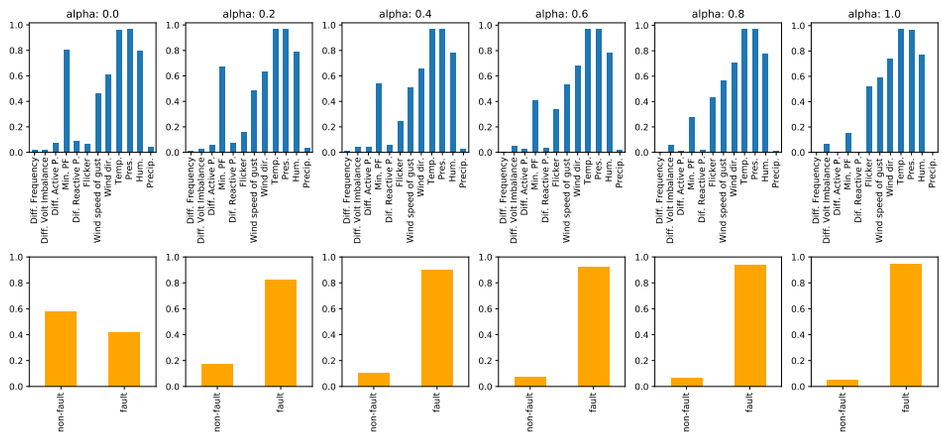


Figure 3.7: Top row: linear interpolation from the mean-baseline (left) to an actual sample of class fault (right). Bottom row: classification probabilities assigned by the MLP at each interpolation step.

False Negative (FN): The number of incorrect observations classified as negative.

To quantify the classification performance from the confusion matrix using one single metric, it is possible to calculate the F1-score, which is computed as:

$$F1 = 2 \cdot \frac{TP}{TP + \frac{FP+FN}{2}}, \tag{3.7}$$

Where a higher F1 score indicates a better classification performance of the model, the highest possible F1 score is 1 (i.e., no FP and FN).

3.3 Optimization of electric power systems

This section presents the background theory of the least-cost optimization techniques used in Paper **IV** and Paper **V**. Least-cost optimization is a widely used strategy in decision-making processes to minimize the cost of achieving a specific objective. Framing a cost-optimization problem involves identifying the most cost-effective way to allocate available resources to achieve a desired outcome. Least-cost optimization is widely used in many applications, such as manufacturing, logistics, and electric power systems (Fazlollahtabar, Saidi-Mehrabad, & Balakrishnan, 2015; Hezam & Nayeem, 2020; Sepulveda et al., 2018, 2021). An optimization problem typically has the following elements:

- Objective function
- Decision variables
- Constraints

The objective function states the problem that must be solved, where the decision variables are the unknown variables that must be optimized to solve the objective function. The decision variables are subject to (s.t) constraints that define the feasible solutions to the system. That is, the decision variables must be within the given constraints to obtain a feasible solution to the optimization problem. A general minimization optimization problem (P) can be mathematically formulated as follows:

$$\begin{aligned} (P) : \min \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{x} \epsilon X \end{aligned} \tag{3.8}$$

where $f(\mathbf{x})$ is an objective function that depends on decision variables $\mathbf{x} = (x_1 ... x_n)^T$. The set $\mathbf{X}$ defines the feasible solutions to the problem. The feasible solutions $\mathbf{X}$ are expressed by constraints, which give the alternative formulation of the optimization problem (P) as

$$\begin{aligned} (P) : \min \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & g_i(\mathbf{x}) \leq b_i, \\ & i = 1, ..., m \end{aligned} \tag{3.9}$$

where $g_1(\mathbf{x}), ..., g_m(\mathbf{x})$ are functions that depends on $\mathbf{x}$, and $b_1, ..., b_m$ are given parameters.

In this thesis, we apply least-cost optimization techniques to analyze and minimize the costs of electric power systems. The least-cost optimization of electric power systems aims to minimize the total cost of generating, transmitting, and distributing electricity while meeting the demand for electricity and satisfying various system constraints. In Paper **IV** and **V**, we employed the least-cost optimization model GenX developed by (Jenkins & Sepulveda, 2017). The GenX model is a CEM which takes the perspective of a centralized power system planner to determine the cost-optimal generation portfolio, energy storage, and transmission investments required to meet a predefined system demand. The objective function with the associated constraints is presented in Tab. 3.2 as:

The full mathematical description of the GenX CEM model is given in detail in (Jenkins & Sepulveda, 2017).

Table 3.2: The GenX optimization problem

Objective:	Minimize total cost over one year (annuitized investment costs + operating costs + penalties for non-served energy and unmet reserves)
Subject to:	
	Demand = Resource Operation (+/-) + Non-Served Energy + Power Flows (+/-) Thermal plant operating constraints (Ramping Limits, Unit Commitment) Electric power generator capacity limits Storage inter-temporal state-of-charge balance Minimum operating reserves requirements (Regulation, Spinning up, down) CO_2 emission limits and/or VRE mandates Capacity reserve margin requirements

Part II

Summary of research

4

Summary of papers

Paper I - Predicting Energy Demand in Semi-Remote Arctic Locations

Forecasting energy demand accurately is essential for developing strategies to manage and optimize available energy resources and the associated infrastructure. This paper presents a framework for selecting the appropriate approach to predict electricity demand accurately at two different locations (location 1 and location 2). These locations represent two remote communities at the end of a radial distribution network without an alternative electricity supply. It is, therefore, crucial to develop an accurate forecasting framework of the electricity demand to manage and optimize the limited electricity resources available. The prediction approaches were divided into ML and statistical methods. The different approaches have advantages and disadvantages, highlighted in Fig. 4.1.

The prediction accuracy was compared on several horizons (from 1 to 165 hours ahead) to gain insights into which approach gave the highest accuracy on different horizons. The ML-based methods provided the most accurate prediction results on all horizons, except on a 24-hour prediction horizon, where an Autoregressive Integrated Moving Average model gained slightly better accuracy. The idea of using ML and statistical models to predict electricity demand time series is not new, and a large body of literature exists on this research. However, the proposed framework provides valuable insight into effectively selecting the most appropriate model when predicting at different time horizons.

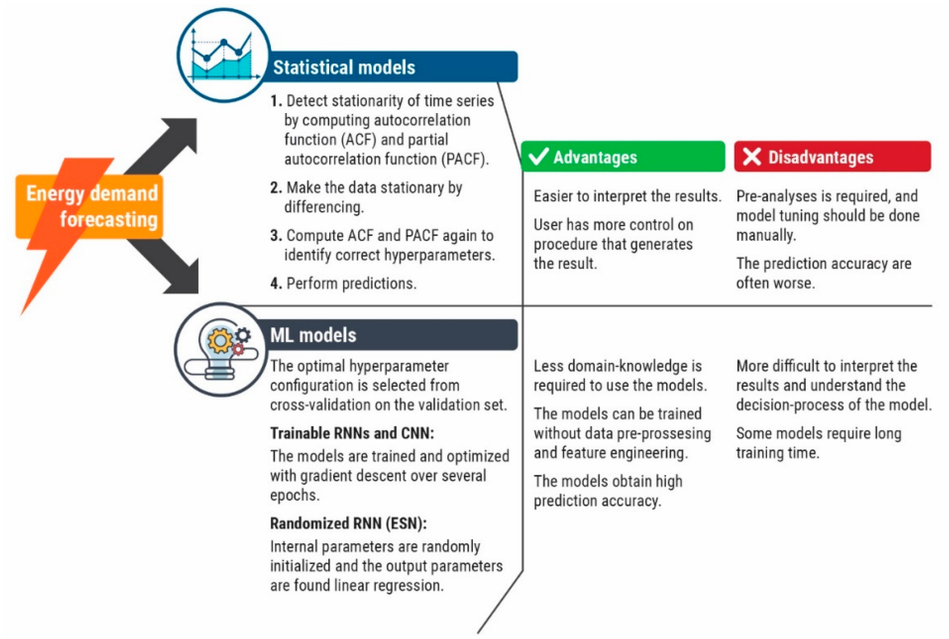


Figure 4.1: Overview of the different approaches to predict the electricity demand. Each method has its advantages and disadvantages.

In addition, this study investigated each method's transferability. Transferability predictions are to train a model on one-time series and use the trained model to predict another time series that are not accounted for during training. This is of value when predicting time series where historical data may be limited or unavailable. To investigate transferability, we optimize the prediction model on the training and validation set of a source dataset (location 1). Then, the trained model is used to predict the test set of the target dataset (location 2). The ML models exhibit good transferability when performing transferability predictions. They managed to predict the load at new locations not accounted for during training with acceptable accuracy. The statistical methods were not useful for transferability predictions and could only be used when trained on the same time series being predicted. Our work will guide in selecting and applying the appropriate prediction model to perform energy load forecasting of different prediction horizons in rural areas and locations where historical consumption data may be limited or unavailable.

Contributions by the author. The idea was conceived by myself and further developed in collaboration with the co-authors. The implementation and experiments were carried out by myself with the help of Matteo Chiesa and Filippo Maria Bianchi. I wrote the draft of the manuscript.

Paper II - Detecting and Interpreting Faults in Vulnerable Power Grids With Machine Learning

This paper was based on two preliminary works. The first was Paper 6, where we spent significant effort constructing the dataset. The second work was a conference paper (Paper 8) presented at the International Conference on Intelligent Technologies and Applications (INTAP) in Grimstad, Norway, in 2021.

In this paper, we predict and interpret the causes of faults in the power distribution networks, which severely affect customers and grid operators. In this work, we focus on the power grid of a Norwegian community in the Arctic that experiences several faults whose sources are unknown, and it is of critical importance to identify these sources to implement strategies to prevent and mitigate incoming faults. We adopt linear and non-linear ML classification models to detect fault occurrences. The experimental results show that the linear and non-linear classification models perform well in predicting faults. The good classification score of all models indicates that the features considered in the dataset explain well the power faults.

To identify the causes of interruptions, it is necessary to interpret the decision process of the ML classification models. First, we used a traditional feature selection method to identify the variables that mostly explain the fault occurrences in the dataset. However, such global interpretation methods only show which variables, on average, contribute to explaining the causes of power interruptions. It is also interesting to identify reasons for specific faults in the dataset. Therefore, we adopt the recent technique IG, to interpret the decision process of a MLP. Fig. 4.2 shows the result from the IG method for one fault sample in the dataset.

The top-left plot corresponds to an "average" sample in the dataset (i.e., uninformative baseline). The blue bar plots represent the value of the features in the selected samples. The green and red bar plots are the output of the IG procedure.

The green bars denote that a feature is important for the classification result. The higher the green bar, the more the feature value in the sample (blue bar) explains the classification result compared to the value in the baseline (black bar). The red bars mean that the value of the features in the sample decreases the classifier's confidence that the sample is a fault occurrence.

The IG approach allows us to understand more deeply what features were important to classify a specific sample as a fault. This is essential when implementing programs to prevent and mitigate potential faults in the distribution

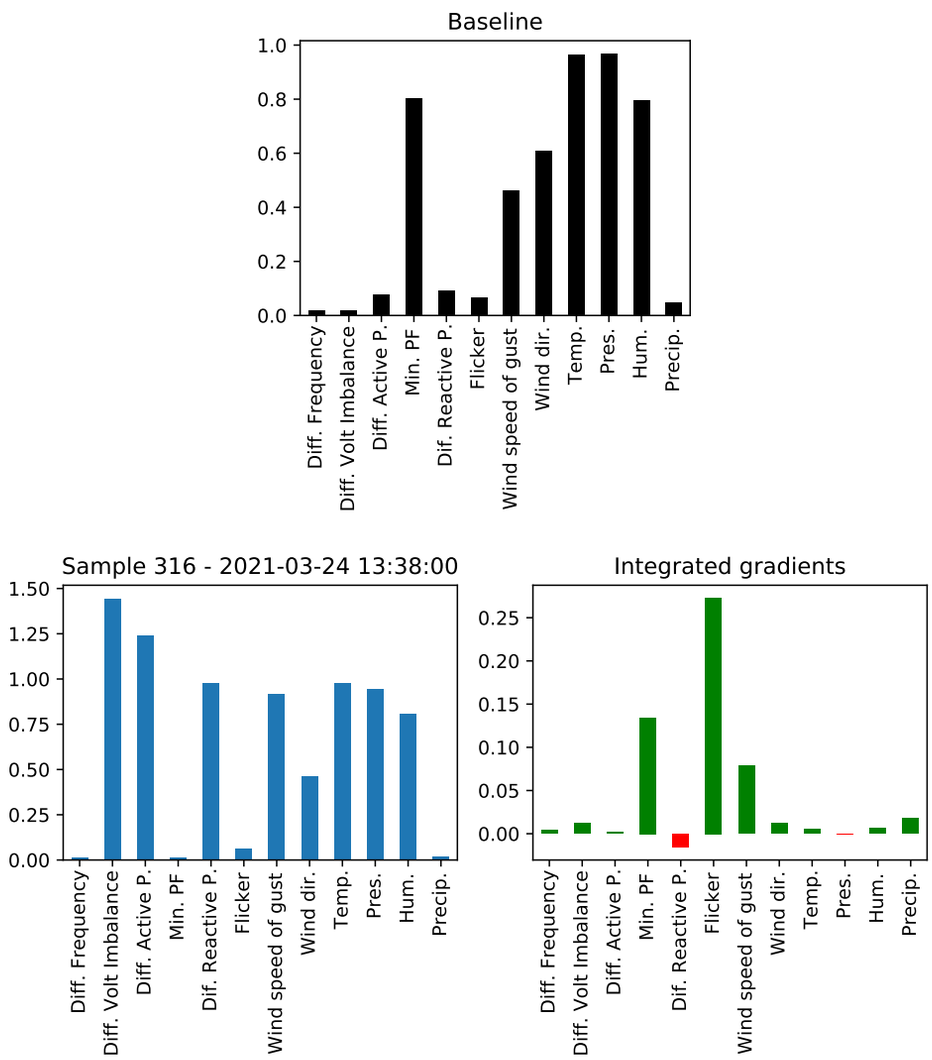


Figure 4.2: The green bars denote that a feature is important for the classification result. The higher the green bar, the more the feature value in the sample (blue bar) explains the classification result compared to the value in the baseline (black bar). The red bars mean that the value of the features in the sample decreases the classifier's confidence that the sample is actually a fault.

grid.

Contributions by the author. The idea was conceived by myself and further developed in collaboration with the co-authors. The implementation and experiments were carried out by myself with the help of Matteo Chiesa and Filippo Maria Bianchi. I wrote the draft of the manuscript.

Paper III - Probabilistic forecasts of wind power generation in regions with complex topography using deep learning methods: An Arctic case

Making accurate predictions about future electricity generation is of fundamental value for the energy market that relies on accurate forecasting capabilities. In this study, we made day-ahead predictions of the electricity generation from a wind power plant in Northern Norway that lies in a region of complex topography and has a highly intermittent nature in electricity generation, increasing uncertainty about future power generation. We computed probabilistic forecasts of expected future electricity generation to account for such uncertainties. When making probabilistic forecasts, the goal is to make a Prediction Interval (PI) that considers the uncertainties in the predictions. The PI estimates an interval where the future observation will fall with a certain probability. For instance, for a 95% PI, there should be a 95% probability that the next value will fall within the lower and upper bounds. A PI can be high and low quality. A high-quality PI is both sharp to convey useful information about the uncertainty and should be calibrated, which means that for a 95% PI, 95% of the observed values should fall within the PI. Fig. 4.3 shows examples of PIs of both high and low quality.

Electricity generation from wind power is directly dependent on the external weather factors. Thus, to optimize the accuracy of the forecasts, it is of fundamental importance to understand which variables (covariates) to include in the prediction model.

We compared the performance of the day-ahead probabilistic forecasts of deep learning models with different sets of covariates. Three different dataset configurations were compared. The first configuration used measured weather data and Numerical Weather Predictions (NWP) on wind speed and direction to predict day-ahead electricity generation. The second configuration used only measured weather data, while the third configuration used only NWP as an exogenous variable in the deep learning model. The configuration where measured weather and the NWP was included gave the highest prediction performance and improved the accuracy by 37% compared to the third configuration when only NWP were used as exogenous variables. The reason is that when predicting the day-ahead weather using NWP, the NWPs often incorrectly estimate the amount of wind since it has a very intermittent and non-linear nature. Using historical measured data in addition to NWPs allows the deep learning model to auto-correct systematic biases in the NWPs. This study shows the importance of gaining insights into which variables mostly affect wind power electricity generation (wind speed and wind direction). In addition, this study highlights

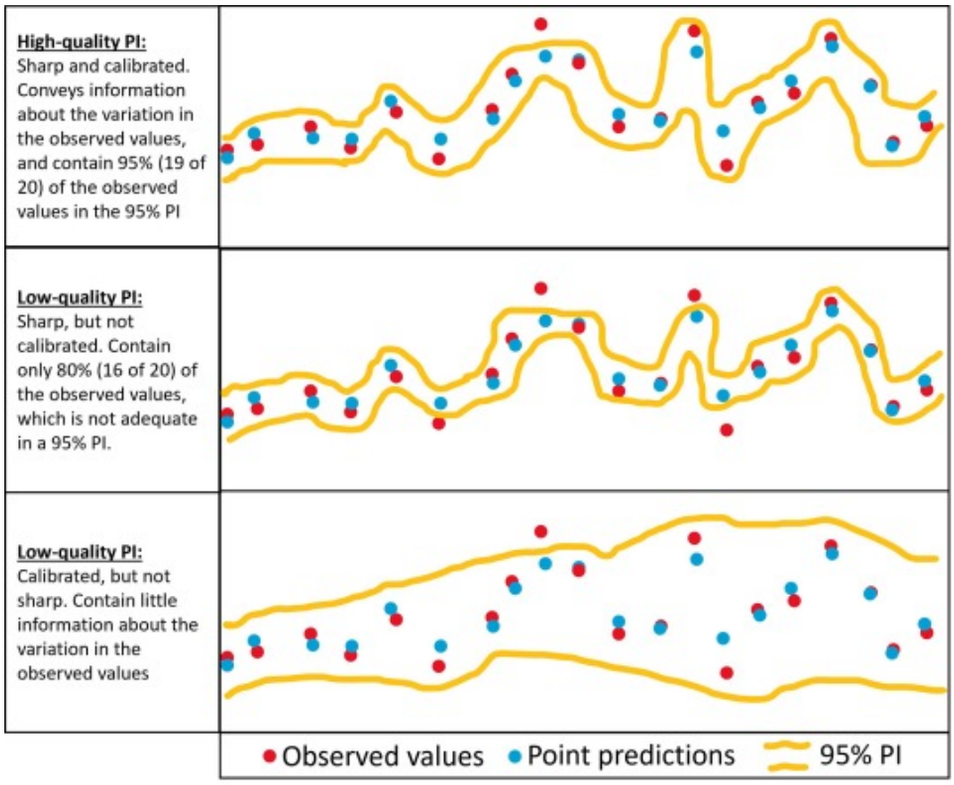


Figure 4.3: Examples of PI with high (uppermost figure) and low quality. The red dots represent the observed values. The blue dots are the point predictions. The yellow line represents the PI, which in this case is the 95% prediction interval.

the importance of improving the accuracy of the NWP weather forecasts.

Contributions by the author. The idea was conceived by myself and further developed in collaboration with the co-authors. The implementation and experiments were carried out by myself with the help of Matteo Chiesa and Filippo Maria Bianchi. I wrote the draft of the manuscript.

Paper IV - Power Availability of PV plus Thermal Batteries In Real-World Electric Power Grids

As variable renewable energy sources comprise a growing share of total electricity generation, energy storage technologies are becoming increasingly critical for balancing energy generation and demand. A major problem of the current state-of-the-art battery technologies such as li-ion is the too-high capital cost between US \$80 $kWh^{-}1$ and US \$100 $kWh^{-}1$ (Braff et al., 2016; Ziegler et al., 2019; Mallapragada et al., 2020). Studies suggest that achieving cost-efficient storage requires a capital cost reduction between US \$3 $kWh^{-}1$ and US \$3 $kWh^{-}1$ (Ziegler et al., 2019; Albertus et al., 2020) to enable full decarbonization of the grid. There exist storage concepts with sufficiently low capital cost, but these are geographically constrained (PHS and CAES). In this work, we addressed these issues by modeling the novel TEGS concept that is both geographically unconstrained and has estimated capital costs that are sufficiently low to enable large-scale deployment in the electric power system (Amy et al., 2019).

We used a CEM to model an existing electricity system representing the New England grid region in Northern America. To analyze the importance of using storage to balance the electricity generation of renewables, we introduced a hypothetical solar energy + TEGS system to the modeled electricity grid. The solar energy system had a capacity of 100 MW, while the TEGS unit has storage capacities between 400 MWh and 600 MWh. The Power Availability Factor (PAF) metric was introduced to investigate the percentage of time during the year the grid could achieve a certain derated amount of power from the modeled solar + TEGS system. The PAF at different derate levels are given in Fig. 4.4.

The modeling was performed under a baseline case with no emission constraints and under hypothetical scenarios where CO_2 emissions were reduced. The results show that the power available to the grid from our hypothetical solar + TEGS system increases when the CO_2 emissions are reduced.

In the CO_2 emission reduction case, the increased retirement of fossil fuel technologies, such as Natural gas, makes the grid more dependent on the hypothetical solar + TEGS system. This results in the system supplying the necessary power 100% of the time for a derated power between 5 MW and 20 MW for the TEGS unit with a storage capacity of 600 MWh and 800 MWh. This is remarkably higher than the baseline case, where the system cannot deliver the required power 5-15% of the time for such derated powers.

The proposed approach provides insight into how adding solar + emerging

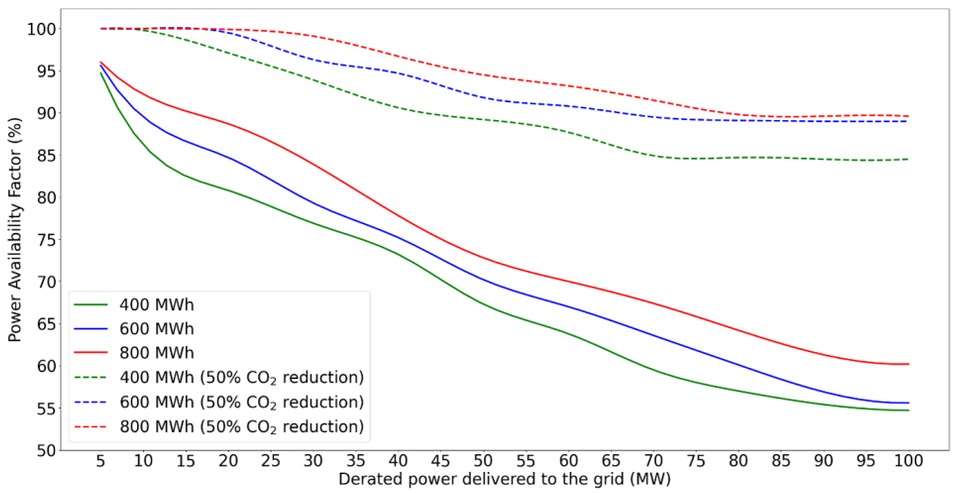


Figure 4.4: Percentage of time during the year when the hypothetical system can deliver the requested power to the grid

storage systems to electric grids can contribute to the efficient stepwise decarbonization of electric power systems by providing more reliable power to the grid on demand.

Contributions by the author The idea was conceived by myself and further developed in collaboration with the co-authors. The implementation and experiments were carried out by myself with the help of Asegun Henry and Matteo Chiesa. I wrote the draft of the manuscript.

Paper V - Cost-Effective Thermal Energy Grid Storage for Decarbonizing Electric Power Systems

This paper was built on preliminary work in Paper **IV**, where the CEM framework and the TEGS technology were the same.

In this paper, we address the problem of the high cost of the most scalable options for electrical energy storage, limiting the amount of renewable energy that can be incorporated into an energy system without significantly increasing the overall cost. We optimize the engineering design of the TEGS unit with respect to obtaining the highest cost reduction compared with a baseline scenario where TEGS is not an available technology. The engineering parameters considered in the optimization schedule were the daily heat loss (from 1% to 10%), operating temperature (from 1900°C to 2400°C), and the charge-discharge ratio. The optimization procedure was computed for four different

CO_2 reduction scenarios (80%, 90%, 95%, and 99%). The modeled electricity systems represented the New England grid in Massachusetts and the Electric Reliability Council of Texas (ERCOT) grid in Texas. The optimum TEGS design under different CO_2 scenarios for the New England grid is presented in Fig. 4.5.

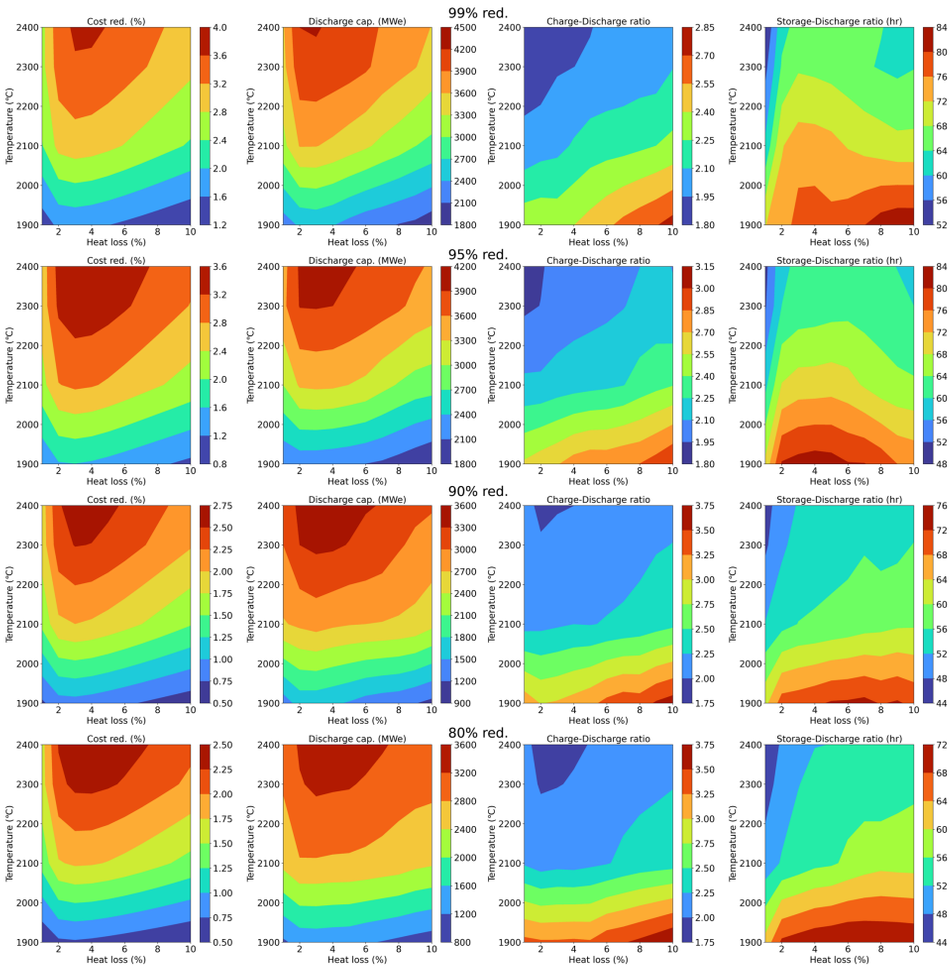


Figure 4.5: Optimum TEGS design under different CO_2 reduction scenarios. For each row, the leftmost figure shows the cost reduction compared to the baseline scenario, and the next figure to the right shows the amount of charging capacity needed. The third figure shows the charge/discharge capacity ratio. The rightmost figure shows the number of hours of storage required.

The results show that the cost-optimized TEGS has a working temperature of 2400°C and a daily heat loss of approximately 3%. This design is the most cost-optimum one for all CO_2 reduction scenarios, and it reduces the cost by approximately 4% in the 99% reduction scenario. To enable such cost-reduction,

approximately 4000 MW of discharging and 55 hours of storage is required. Interestingly, the charge-discharge capacity shows that the TEGS unit should be capable of charging twice as fast as discharging to charge a large amount of energy when the electricity demand is low. Similar results were obtained when modeling the ERCOT grid.

In addition to enabling cost-efficient decarbonization of the electricity system, it is critically important that storage units contribute to ensuring the grid's reliability by serving the grid with sufficient power when there is a lack of electricity generation from renewable energy sources. The annualized Non-Served Energy (NSE) cost was computed to measure resiliency. The NSE cost is computed as the number of hours during the year when the demand is unmet, times the value of the lost load. Fig. 4.6 show the cost for NSE for the baseline and optimum TEGS scenario over 22 years under 99% CO_2 reduction in the New England grid.

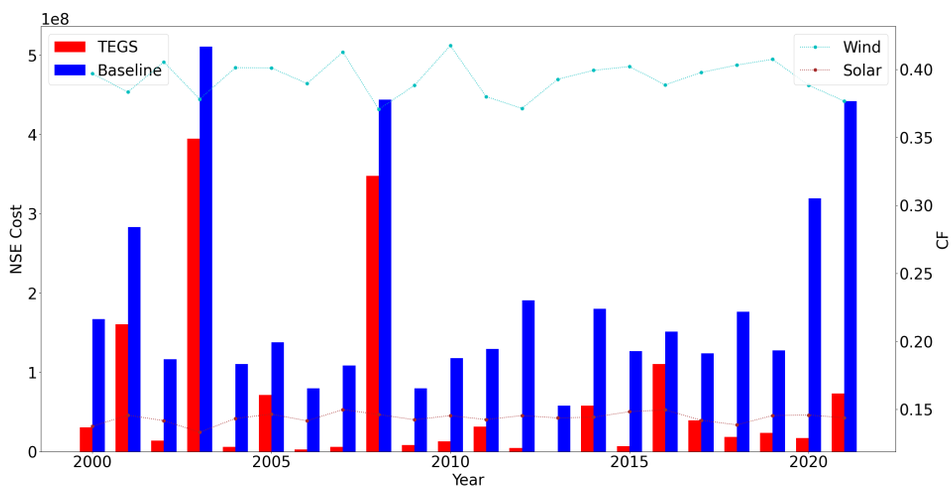


Figure 4.6: Cost of non-served energy for weather years with different VRE availability

It is clear that TEGS vastly reduces the cost of NSE compared to the baseline case due to the larger amount of available storage technologies that can serve the grid when there is a lack of solar and wind availability. The findings show that design-optimized storage units can be essential in obtaining cost-efficient decarbonization of electric power systems while maintaining resiliency.

Contributions by the author. The idea was conceived by myself and further developed in collaboration with Ruaridh Macdonald. The implementation and experiments were carried out together with Ruaridh Macdonald under the supervision of Asegun Henry and Matteo Chiesa. I wrote the draft of the manuscript.

/5

Concluding remarks

In this thesis, we have addressed some important challenges, goals and objectives regarding the operation and decarbonization of the energy sector. Our solution to address these goals was to employ machine learning and optimization techniques to gain deep insight into energy-related datasets. In the first and third works, we presented frameworks to predict the time series of energy datasets by comparing several statistical and machine-learning-based models. We found that constructing the datasets and defining the correct sets of covariates is essential to improve accuracy when the time series are highly nonlinear.

We also analyzed the problem of detecting the distribution network's power faults whose sources are unknown. In the second paper, we used machine learning classification models to predict the occurrence of faults in a distribution network that serves a Norwegian community in the Arctic that experiences several faults whose sources are unknown. A significant focus of this work was to gain insight into the problem by collaborating with industry experts and thus constructing a dataset with variables that could explain the fault occurrences. The resulting dataset consisted of variables divided into weather-related and power-quality-related variables. The machine learning classification models performed well when predicting the fault occurrence, indicating that power quality and weather variables explained the power disturbance well. To prevent and mitigate power fault occurrences, the second work also adopted a machine learning interpretation technique to identify and understand the main causes of faults in the electricity system.

In addition to using machine learning methods, this thesis has presented two works where optimization techniques have been used to model electricity systems. The optimization was done to gain insight into the importance of using emerging storage technologies to reduce the cost of decarbonized grids and maintain reliability by balancing the intermittent electricity generation from renewables. In Paper **IV**, we modeled a hypothetical solar + TEGS system and found that the TEGS storage concept contributed to improving grid reliability by discharging power when there is a lack of solar availability. In addition, the importance of using such combined solar + TEGS systems increases in the modeled CO_2 reduction scenarios as the grid must be less dependent on fossil-fuel-based technologies. This thesis's fifth and final work assessed the importance of using emerging storage technologies to enable cost-efficient decarbonization. We cost-optimized the engineering design of the TEGS unit and found that such technologies can reduce the grid's overall cost by approximately 4% when compared to a baseline scenario. In addition, we found a large benefit of using such storage units to balance renewable energy generation as it significantly improves the grid's reliability by discharging power on demand. This work provides important insights into optimizing storage for cost-effective grid decarbonization while maintaining resiliency.

We conclude that, with the five works of research presented in this thesis, we contributed to advancing the field of energy analytics, mainly by addressing the objectives of Improved performance and reliability, renewable energy integration, decarbonization, and cost savings.

5.1 Limitations and further work

We acknowledge that every research paper has both strengths and weaknesses. Therefore, we end the concluding remarks by discussing limitations, practical applications, and suggested future work for the research presented in this thesis.

Paper I. We mentioned in Paper I that interesting future work is to investigate the possibility of combining the Echo State Network (ESN) and CNN models for long- and short-term prediction purposes. If such a model could achieve high prediction results on multiple time series on both long and short-term horizons, it would be a valuable prediction tool for energy planning purposes at several locations and sectors with different time series dynamics. We have also identified other weaknesses that should be considered in future work.

All models in this paper performed univariate predictions where we trained the models on only one time series considering the historical data of electricity de-

mand. The electricity demand is affected by external weather factors (i.e., cold weather increases the demand for electricity due to heating). Suggested future research is to consider the same time series, but do multivariate predictions to consider external factors that might explain the future electricity demand. However, given the high prediction accuracy of this work, it is questionable if such multivariate predictions will significantly improve the accuracy.

In addition, this study covers two locations that are in an Arctic environment with harsh weather conditions. It would be interesting to repeat the suggested framework for this paper in other locations with different climates. If our framework could be applied to multiple locations, it would be useful in different practical applications, such as contributing to setting the electricity price day ahead by bidding the expected demand for electricity into the market.

Paper II. The dataset construction methodology was a major part of this paper, built on the preliminary works (Paper 6 and Conference Paper 8). In collaboration with expert knowledge within the DSO, we decided to collect weather data from weather-exposed areas in the grid to detect possible weather-related failures in the grid. Two assumptions were made to collect the weather data from these areas: One was to assume that higher elevation increases the probability of exposure to harsh weather conditions, such as strong wind. Indeed, utility poles at high altitudes are often in mountainous areas where there is no vegetation that can protect from the wind. However, this approach neglects other grid areas that might contribute to explaining possible weather-related fault occurrences. There have already been efforts building on our work where the authors have applied a more theoretical methodology to collect the weather data that can explain fault occurrences. However, the authors ended up with similar results as in this study, which shows that our approach to identifying the weather-exposed areas is useful when collecting data that might explain the weather-related fault occurrences.

This study covers an Arctic region with harsh weather conditions, and there is not much vegetation that can fall on the transmission lines and thus result in power interruptions. Therefore, our dataset, which does not include vegetation information, will not be very useful in other regions with other climates which have problems with trees falling on the power lines, as analyzed in the study in (Gazzea, Aalhus, Kristensen, Ozguven, & Arghandeh, 2021).

However, this study has several useful, practical applications for DSOs. If the DSOs understand the sources of potential fault occurrences, they can implement specific programs to strengthen the grid and thus avoid incoming faults. Potential actions to improve the grid stability are: i) make changes in grid topology, such as optimizing coupling to make the grid stronger, isolating parts of the grid more likely to fail, running island mode whenever possible; ii) optimizing

or even increasing the local production by introducing new power sources, including renewable ones; iii) reduce or adjust power flows by controlling flexible loads.

These kinds of strategies to mitigate incoming grid faults are currently being developed by the DSO operating the grid in this study. In particular, the local power company has installed a battery system that should be activated before an incoming power fault. Using the methodology in this study to understand which variable should be monitored to detect an incoming power fault is fundamental to optimizing the operation of the batteries. An interesting future study is to analyze how many potential faults the batteries can prevent by knowing which variables likely will cause a power fault.

Paper III. The wind power farm analyzed in this work consisted of 18 different wind turbines, and we predicted the aggregated electricity generation from all the turbines. However, during the pre-analyses of the wind farm, we noticed differences in the electricity generation profiles between each turbine due to local variations in wind speed and wind direction. When aggregating the electricity generation from all individual turbines, the individual differences in the generation profile are consequently neglected. Therefore, a future interesting study is to develop a forecasting framework where a ML model is trained and optimized for each turbine's generation profile. This ensures a model that will consider local variations in the wind power generation profile and thus might improve the overall prediction accuracy. However, as a downside, this will significantly increase the computational intensity as the ML model must be trained and optimized on the time series of 18 different wind turbines.

This work proposed a prediction approach to predict future electricity generation and accounted for the uncertainties in the predictions by making probabilistic forecasts. However, since RE technologies participate in the electricity market, it is interesting to address the potential benefit of improving prediction accuracy in terms of reducing the potential financial penalties when mismatches between contracted generation and actual deliveries occur. Similar to the work by (Mazzi & Pinson, 2017), a future suggested study is to incorporate our probabilistic forecasting approach into a financial model and compute the reduction in financial penalties as a function of improved prediction accuracy.

Paper IV. This study presented an idealized representation of an existing grid in Northern America. However, the grid representation might not fully capture all details of the existing grid. There can be differences (sizes of the power plants, electricity demand on the grid, share of the existing generation technologies) between the abstract grid representation and the current real-world grid. In addition to the transmission line between the existing grid and the hypothetical

solar + storage system, the current grid was modeled as a single-zone grid region without considering transmission losses or congestion between generators and demand.

The power availability of the solar + TEGS system changes significantly as a function of solar availability. Therefore, testing the PAF at other locations with different solar availability would be interesting. Additionally, we modeled one year of weather data. It would be interesting to test the PAF under different weather conditions by modeling several weather years. Also, we made assumptions about the technical specifications of TEGS by assuming a 1% heat loss and a 2400°C working temperature. This is not necessarily the most cost-efficient design. In the next paper (Paper **V**), we considered this by optimizing the engineering design that should preferably be used in the grid.

Paper V. In this work, we modeled representations of two electricity systems in Northern America using a "greenfield" approach (i.e., starting from scratch). Consequently, we ignore the existing electricity system infrastructure and modeled fully hypothetical power systems with respect to how to obtain the most-cost efficient grids under different decarbonization scenarios. As we were interested in modeling the value of using storage in hypothetical decarbonized grids, we modeled the grids as a single-zone region without considering transmission losses or congestion between generators and demand as it was outside the scope of the study and will significantly increase the computational intensity of the CEM.

Additionally, we made a significant effort to optimize the engineering design of TEGS to minimize the system's cost. However, due to the significantly increased computational complexity of the CEM when modeling non-linearities, we assumed linear charging/discharging capacities and heat losses. In reality, the exact operation of the TEGS system has non-linear behavior. Consequently, the findings in this paper might not represent the exact cost benefit of utilizing TEGS. However, after discussions with the lab designing the TEGS technology, the authors have reasons to believe that the linear assumptions represent the TEGS operation mechanism well.

Part III

Included papers

/6

Paper I

energies

MDPI

Article

Predicting Energy Demand in Semi-Remote Arctic Locations

Odin Foldvik Eikeland [1], Filippo Maria Bianchi [2], Harry Apostoleris [3], Morten Hansen [4], Yu-Cheng Chiou [1] and Matteo Chiesa [1,2,*]

[1] Department of Physics and Technology, UiT the Arctic University of Norway, 9037 Tromsø, Norway; odin.f.eikeland@uit.no (O.F.E.); yu.cheng.chiou@uit.no (Y.-C.C.)
[2] Department of Mathematics and Statistics and NORCE, The Norwegian Research Centre, UiT the Arctic University of Norway, 9037 Tromsø, Norway; filippo.m.bianchi@uit.no
[3] Laboratory for Energy and NanoScience (LENS), Masdar Institute Campus, Khalifa University of Science and Technology, 127788 Abu Dhabi, United Arab Emirates; harry.apostoleris@gmail.com
[4] Ishavskraft Power Company, 9024 Tromsø, Norway; morten.hansen@ishavskraft.no
* Correspondence: matteo.chiesa@uit.no

Abstract: Forecasting energy demand within a distribution network is essential for developing strategies to manage and optimize available energy resources and the associated infrastructure. In this study, we consider remote communities in the Arctic located at the end of the radial distribution network without alternative energy supply. Therefore, it is crucial to develop an accurate forecasting model to manage and optimize the limited energy resources available. We first compare the accuracy of several models that perform short-and medium-term load forecasts in rural areas, where a single industrial customer dominates the electricity consumption. We consider both statistical methods and machine learning models to predict energy demand. Then, we evaluate the transferability of each method to a geographical rural area different from the one considered for training. Our results indicate that statistical models achieve higher accuracy on longer forecast horizons relative to neural networks, while the machine-learning approaches perform better in predicting load at shorter time intervals. The machine learning models also exhibit good transferability, as they manage to predict well the load at new locations that were not accounted for during training. Our work will serve as a guide for selecting the appropriate prediction model and apply it to perform energy load forecasting in rural areas and in locations where historical consumption data may be limited or even not available.

Keywords: energy load predictions; statistical- and machine-learning-based approaches; short-term load forecasting; longer forecasting horizons; transferability predictions

Citation: Foldvik Eikeland, O.; Bianchi, F.M.; Apostoleris, H.; Hansen, M.; Chiou, Y.-C.; Chiesa, M. Predicting Energy Demand in Semi-Remote Arctic Locations. *Energies* **2021**, *14*, 798. https://doi.org/10.3390/en14040798

Received: 29 December 2020
Accepted: 27 January 2021
Published: 3 February 2021

1. Introduction

Accurate load forecasting systems can reduce additional costs related to inaccurate prediction of the energy demand and provide a better understanding of the dynamics of existing power systems [1,2]. If the forecasts overestimate the demand, the result will be excess power supply. Consequently, this will result in increased costs and contract curtailments for the energy market participants. On the other hand, underestimation of the demand could lead to a lack of energy availability at heavy loads, which in turn leads to consequences for end-users, who in the worst-case scenario will not have sufficient energy supply [3,4]. The global energy market is now evolving from centralized systems with large power stations connected to a single electricity grid which support the area of interest, towards the inclusion of more decentralized energy systems where the area of interest may be supplied by multiple energy sources, such as local renewable distributed generation (DG) technologies and battery storage systems [5–11]. In addition, inhabitants should have the ability to participate actively in the energy market by acting as prosumers where they both generate and consume electricity [12–17].

In the transition from centralized to decentralized energy systems, there is a need to develop methodologies for understanding and interpreting the dynamics of existing energy systems, which in turn will be an essential tool in developing emerging energy systems [18–20]. The Norwegian Water Resources and Energy Directorate (NVE) states that using the current electricity grid costs the households and industries in Norway approximately 27 billion NOK per year [21,22]. This cost is expected to increase significantly in the future, as the current electricity network must be upgraded to handle the increased electrification in society. Therefore, NVE has proposed new power tariff schemes that penalize the use of electricity during periods of heavy loads [21]. By optimizing the energy flows in the electricity network, the profile of the load demand will become more evenly distributed throughout the day and could allow postponement of upgrades to the current infrastructure. The requirements by NVE show the importance of optimizing the use of resources in an energy grid and developing accurate energy-forecasting systems are fundamental to achieving this goal.

To manage the limited availability of energy at locations that rely on the stable power supply, accurate predictions of energy demand are essential. In this study, we analyze two remote communities on the island of Senja in Northern Norway, which represent extreme examples of the challenges stated by NVE. The two communities experience frequent power outages due to heavy loads in specific periods and, if the current grid is not optimized, upgrading the whole energy system will be inevitable. The energy consumption on the grid is heavily affected by the activities of the islands' major industries [23].

Besides Northern Norway, the challenges mentioned above also apply to remote communities worldwide, especially developing countries [24,25]. In 2018, there were approximately 860 million inhabitants of remote areas that either do not have reliable sources of electricity or are not connected to power grids at all [26]. In addition, remote areas in Russia, Alaska, and Canada are heavily dependent on diesel generators for electricity supply, which imposes an economic, environmental, and social burden on the local populations as diesel generators create noise pollution, aggravate the local air quality and increase the emissions of greenhouse gases [27]. These examples confirm the importance of transitioning towards new energy systems to improve economic, social, and environmental conditions, in addition to ensuring a reliable power supply [27–29].

In the remote communities served by the Senja network analyzed in this paper, there are still no services available for predicting the energy demand [23]. We evaluate approaches to forecasting energy demand based on both statistical and machine learning-based approaches to project future energy demand from historical data. The total energy demand is characterized by a combination of load profiles from two sectors (households and industry) that exhibit very different consumption profiles and require dedicated forecast models. Unlike energy load profiles in cities and residential communities, where the household sector is the main contributor to the total energy consumption, the total load profiles for the remote communities analyzed here are dominated by industry [30]. Industrial activities are therefore essential to consider when developing an accurate forecasting model.

The contribution of our work is twofold. First, we investigate which model achieves the best performance in predicting the energy load in rural areas, as a function of the forecast horizon. Our analysis evaluates several statistical and machine learning approaches.

Secondly, we analyze the transferability of each prediction model, in terms of the capability of predicting time series of energy demand at different locations within the Senja electricity grid. Our study provides important insights about the possibility of applying the models considered in our study to new geographical sites.

2. Background and Related Work

The energy load profiles are typically represented by time series that describe the dynamics of the underlying energy distribution system and are characterized by typical human-based seasonal and cyclic consumption patterns. Indeed, load patterns can vary

significantly depending on the observation period, the nature of each system, and external factors such as climate and weather. Achieving accurate load forecasting has been one of the principal foci in several research areas. Different models have been proposed, each one characterized by different properties in terms of complexity and effectiveness in predicting on different forecasting horizons [31–33].

In this paper, we examine several approaches to predict energy consumption on both short and longer forecasting horizons. Two different forecasting methodologies are used, one statistical-based method, and one method based on neural networks.

2.1. Statistical Methods

For the statistical method, two different approaches were investigated. First, we consider the statistical Autoregressive Integrated Moving Average (ARIMA) model, which has been a baseline tool in prediction problems for several years [34]. The ARIMA model is well known for its simplicity of implementation and high accuracy in predicting long forecasting horizons [35]. The second approach is a newer statistical model called Prophet. Prophet is a tool that has recently gained popularity due to its simplicity and flexibility when performing predictions [36].

ARIMA models require an elaborated statistical analysis to optimize the model configuration [37]. Moreover, ARIMA models make strong assumptions about the nature of the underlying dynamical system that generates the observed time series. This usually introduces strong biases that, in some cases, might be wrong and will hinder the accuracy of the predictions [38]. In the study by Taylor [35], the ARIMA predictions at longer forecasting horizons were shown to outperform more complex models in terms of accuracy. In Reference [39], short-term load forecasting with window-based ARIMA algorithms was applied to predict electricity consumption to optimize the available energy supply. The study concluded that it is sufficient to consider daily consumption data and aggregated hourly coefficients of daily profiles to obtain accurate short-term predictions by use of the sliding window-based forecasting algorithms proposed by the authors. In the study by Bianchi et al. [40], the authors proposed a comparative study for heat demand forecasting in a real-world case. The authors tested ARIMA models on different load time series and the results showed that the ARIMA models can perform accurate predictions on long-term horizons.

Prophet is implemented as an open-source library designed for making predictions on univariate time series [36]. The library is easy to use and allows the identification of optimum hyperparameter configurations for the model that will make a forecast of the time series. The Prophet library offers a practical prediction tool that can be used by analysts without expertise in time series modeling. In the original paper [36], the authors compare the Prophet method against several automated forecast procedures such as ARIMA, exponential smoothing models, random walk model with weekly seasonality, and a TBATS model with both weekly and yearly seasonality. The result shows that the Prophet forecasts result in lower prediction errors.

2.2. Machine Learning Methods

Five different neural network architectures are considered to predict the energy demand. In particular, we examine both Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs).

In prediction-related problems, RNNs have gained significant attention due to their ability to capture complex non-linear dynamics in the time series [41], and RNN architectures have been shown to outperform other models in tasks related to forecasting energy demand [42]. The RNNs examined in this paper have been previously applied to solve different types of prediction related tasks [43]. As trainable RNN architectures, we consider Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU), and Elman Recurrent Neural Networks (ERNN). In addition, we consider the Echo State Network (ESN), which is a randomized neural network from the family of Reservoir Computing approaches [42].

The RNNs are capable of modeling any non-linear dynamical systems up to a given precision and have been applied in many contexts where the temporal dependency is an important feature that characterize the data [41]. A popular RNN architecture is the LSTM, which provides the capability of storing information for long periods of time [44,45]. In load forecasting problems characterized by complex and variable temporal dependencies, RNNs have been shown to outperform several forecasting methods like ARIMA, Support Vector Machines (SVM), Multilayer Perceptron, and Adaptive Network-Based Fuzzy Inference System (ANFIS) [40]. The authors in Reference [43] reviewed and evaluated two real-world datasets of electricity load, comparing modern deep learning architectures on short-term load forecasting problems. In particular, the paper focused on feedforward and recurrent neural networks, sequence-to-sequence models, and temporal convolutional neural networks. The authors found that on short-term load forecasting problems, the simpler ERNN performs comparably to more advanced networks such as GRU and LSTM when adopted in aggregated load forecasting cases. Therefore, in such cases, the authors conclude that the ERNN may represent the most effective solution as it offers the benefits of a low-complexity model without compromising the prediction accuracy.

The Echo State Network is an RNN model that has received increased interest among researchers for simplicity of training and high-accuracy performance in predicting real-valued time series [42]. In predictions of real-world energy load time series, the application of the ESN model has achieved state-of-the-art results on 1-h and 24-h forecast horizons [31,40,46–48]. The authors in References [31,49] performed predictions by using ESN and ARIMA approaches and combined the models to obtain high-accuracy results. In a study by Jaeger and Haas [50], the ESN model was applied on both real and benchmark datasets, and the authors highlighted the capability of the network to compute accurate predictions even when forecasting chaotic processes. In a comparative study by Bianchi et.al [40], the ESN model was tested against several RNNs on synthetic and real-world time series in different contexts. The result showed that the ESN architecture was competitive in most tasks in terms of prediction accuracy. Moreover, the simplicity of its implementation and training compared to other RNNs makes the ESN architecture especially attractive. For load forecasting problems on real-world time series, the ESN architecture was therefore concluded to be a convenient prediction tool for real-valued time series [40].

So far, CNNs have received less attention compared to RNNs in energy load forecasting tasks [43]. However, recent work indicates that CNNs achieve high performance in several sequence and time series prediction tasks [51]. The authors in Reference [52], developed a deep CNN called DeepEnergy. The experimental results show that DeepEnergy can predict energy loads with high accuracy over three days and outperforms SVM and LSTM. The authors of Reference [53] compared a CNN to recurrent and feed-forward architectures, showing promising results on benchmark time-series. Several works combined CNNs together with RNNs to achieve a hybrid prediction tool that increases the prediction accuracy. For example, the authors in Reference [54] combined a CNN and RNN to integrate different input sources and use the convolutional layer to extract features from the historical data. The RNN was thereafter used to learn the dynamics of the system. In Reference [55], another hybrid model was presented where the historical load was processed by a CNN and LSTM. Then, the features from both networks were used to predict the day-ahead load.

2.3. Transfer Learning

A large amount of scientific work on forecasting time series has resulted in a large number of methodologies that provide accurate results in several prediction-related challenges. However, accurate forecasting of time series is a challenging task when the availability of training data is limited. Recently, the transferability of machine learning models has gained increased attention due to the necessity of generating predictions in systems where training data are not available. In Reference [56], the authors investigated how to transfer CNNs for the time series classification (TSC) task. A hybrid transfer learning model for

short-term electric load forecasting was proposed by Reference [57] and shows significant improvement in electric load predictions for a location by using additional data from another location. In Reference [58], the monthly electric load was predicted by collecting data from different districts in Seoul. After collecting all historical data from the different districts, the model was trained to predict the electric load in a target district by using transfer learning. The authors also proposed a novel load-forecasting scheme for a district by using similar data from other cities or districts. To demonstrate the performance, the model was compared to other popular machine learning techniques. The authors in Reference [59] considered the problem of developing predictive models with limited data for energy assets. An energy predictive model based on CNN's was developed and applied by using a transfer learning strategy. The approach was demonstrated on a case study of daily energy demand time series. In Reference [60], the authors show for the first time that using an LSTM auto-encoder with attention trained on a large-scale dataset with pre-processing can effectively transfer time-series features. The authors in Reference [61] proposed a new training strategy for time-series transfer learning with two source datasets that outperforms existing approaches when predicting a target dataset. The authors tested the performance of the approach on predicting financial time series, and the experiments show that transfer learning based on two datasets (market indexes) is superior to other baseline methods with only one source dataset.

While not comprehensive, this short overview gives an idea of the significant amount of research that has been carried out on developing forecasting methodologies. In addition, the transferability of forecasting models has recently gained increased interest, and several works have addressed the problem of accurate predictions when there are limited data available for training.

3. Methodology

In this study, three trainable RNNs (LSTM, GRU, and Elman), a randomized RNN (ESN), a CNN, and two statistical methods (ARIMA and Prophet) are compared for predicting energy demand for multiple cases. The models are trained to predict energy demand at the next hour and at longer time horizons (2, 6, 12, 24, and 165-h ahead). The predictions are performed on two sectors (Households and Industry) at both communities of the Senja grid.

Section 3 explains the process of training the model and how to transfer them to a new dataset. A brief description of the models used (LSTM, GRU, Elman, ESN, CNN, ARIMA, and Prophet) is deferred to the Supplementary Materials (Chapter S7). In addition, the Supplementary Materials briefly discusses the advantages and disadvantages of the different approaches. All hyperparameter configurations used for predictions with the different models are listed in Section S7.5 in the Supplementary Materials Section.

An overview of the methodology for making predictions with the different approaches are provided in Figure 1. A brief explanation of each step is given in the Supplementary Materials Section.

3.1. Prediction Strategy

3.1.1. Training, Validation, and Test

To train and evaluate the models, the time series is split into a training set (70%), validation set (15%), and test set (15%). The training set is used to fit the model parameters by minimizing the prediction loss; the validation set is used to find the optimal configuration of the hyperparameters and to compute the stop criterion for the models trained with gradient descent; once the optimal model is found, its performance is evaluated on the test set. We pre-process the data by removing the linear trend and the main seasonality (see Section 4.2 for more details). Additionally, the data are standardized by subtracting the mean and dividing by the standard deviation computed on the training set.

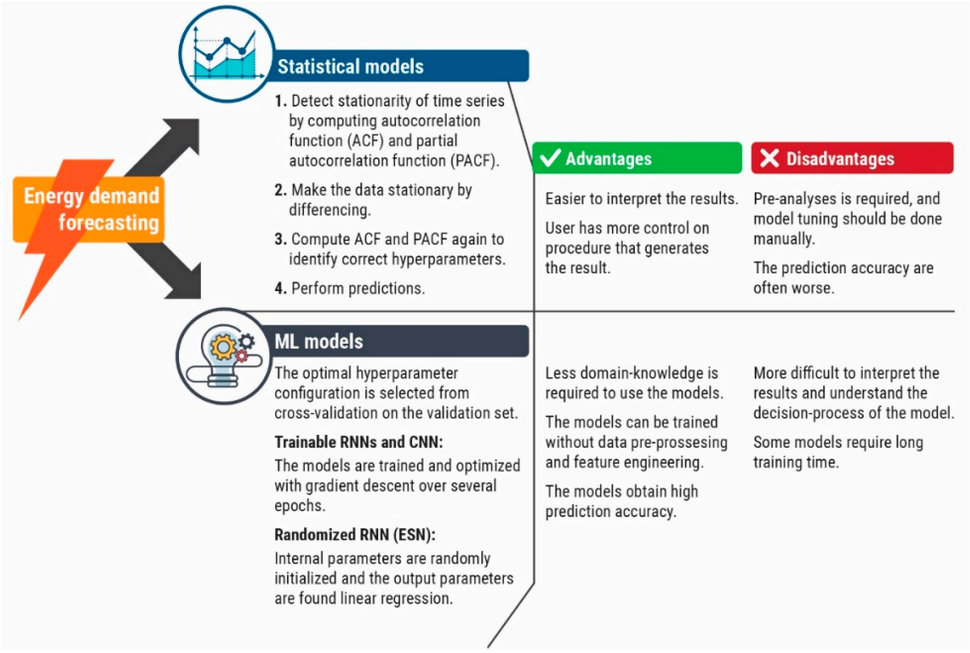


Figure 1. Overview of methodology for making predictions with statistical-and machine-learning models. The advantages and disadvantages of each approach are given.

3.1.2. Transferability

The strategy used to transfer the model is similar to the one proposed in Reference [61]. The model is first optimized on the training and validation set of a source dataset (location 1). Then, the trained model is used to predict the test set of the target dataset (location 2). The data of both source and target datasets are pre-processed according to the statistics computed on the target dataset.

3.2. Normalized Root Mean Squared Error

The normalized root mean squared error (NRMSE) is used both as the loss function and as the metric to evaluate the prediction performance. The root mean squared error (RMSE) is defined as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \quad (1)$$

where $\hat{y}$ and y are the predicted and true values, accordingly. The RMSE measures the discrepancy between the predicted values and observed values at time i, over n number of observations [62]. The NRMSE relates the RMSE to the observed average value in the observation period and is defined as:

$$\mathrm{NRMSE} = \frac{\mathrm{RMSE}}{\bar{y}} \quad (2)$$

where $\bar{y}$ is the average of the time series values.

4. Case Study

At present, the 66 kV cable that supplies Senja is operating close to its maximum capacity. At Northern-Senja, the communities are provided with a 22 kV distribution network. The total energy demand is characterized by a combination of load profiles from two sectors (households and industry) that are very different from each other. The particular feature of the rural communities in Senja is that industry accounts for more than 50% of the total energy consumption [23]. When the industries operate at heavy load, the risk of voltage drop at the end of the radial distribution network increases. Voltage drops increase

the risk of interruption in the power supply. These communities have no alternative energy supplies that can provide a backup when the voltage decreases. Therefore, given the limited capacity of the existing grid, it is paramount to predict the periods with high consumption. An accurate prediction tool is necessary for developing efficient strategies to reduce the challenges in the existing grid. One way to deal with the current situation is to build a new electricity grid connection with a higher capacity. However, this is costly and time-consuming, has a huge environmental impact, and contradicts the vision to better utilize the current electricity grid as stated by NVE [22]. Moreover, since the communities are strongly industry-dominated, such an investment will be wasted if the companies will stop their activities in the future, and the power companies will be left with an over-dimensioned distribution network.

4.1. Real-World Time Series

The time series analysis in this paper is based on hourly energy consumption data in the period spanning from 1 March 2019 to 1 February 2020. The time series are from two separate communities located approximately 9 km apart in air distance, located at the end of separate radials of the electricity network in the region. Both communities are small with approximately 100 households at location 1, and 300 households at location 2.

The time series is provided from the power company Ishavskraft AS [63], which collects consumption data for all their customers in the specific locations studied in this paper. The time series is divided into three sectors: cabins, households, and industry. Due to privacy policies, the household and cabin consumption is aggregated. From Figure 2, it is possible to see that the industry sector in both locations is clearly the dominant source for the total energy consumption pattern and the total energy consumption is significantly larger for location 2 than for location 1, with average hourly consumption of 2553 kW/h and 860 kW/h, respectively. Figure 2 illustrates that the consumption pattern decreases significantly during the main holiday periods (Easter at end of April, summer vacation in July, and Christmas in December/beginning of January). The period between September and December is characterized by an increase in the total energy consumption due to the electric heating because of colder temperatures. In addition, during this period the fishing industries are heading towards the main fishing season when they operate frequently at full load.

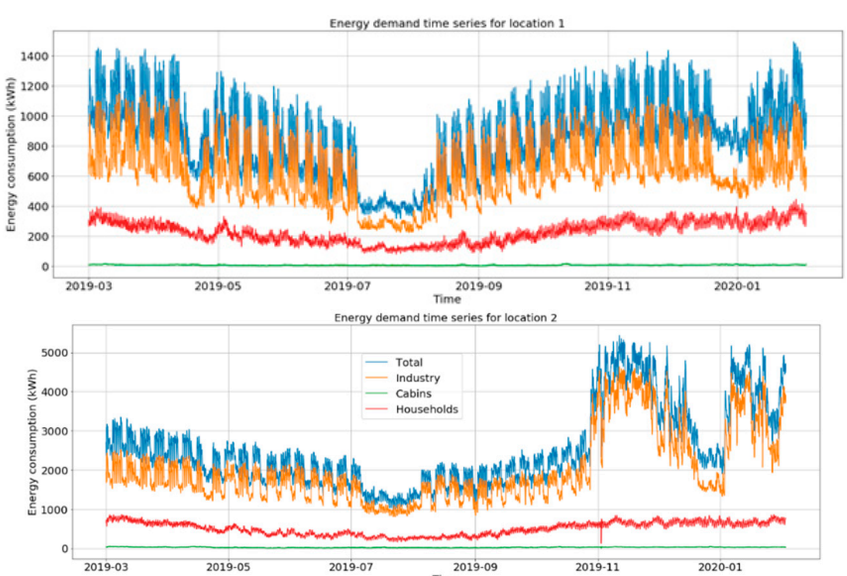


Figure 2. Real-world time series of consumption for location 1 and 2. The demand for the industry sector is strongly correlated with the fishing season, while the energy demand for households is correlated with temperature changes, where cold temperatures during wintertime give higher energy demand for heating purposes.

To design an accurate predictive model, it is necessary to account for the different consumption patterns, which vary between the holidays and the working periods.

We disregard the holiday periods, as the requirements for stable power supply are not as critical as they are during working days. Consequently, we only consider the workdays that are characterized by higher load consumption that is critical given the limited capacity of the grid, and there is a risk of interruptions in the power supply.

For this reason, we train our prediction models only to predict the consumption during workdays. The workdays also include the weekends, because especially in high-season, the fishing industries have a high level of activity every day, including weekends. Specifically, the periods with low consumption during holidays are removed from the data set and we obtain a time series with 6544 time-steps in total. The resulting time series for both locations with holidays removed are depicted in Figure 3.

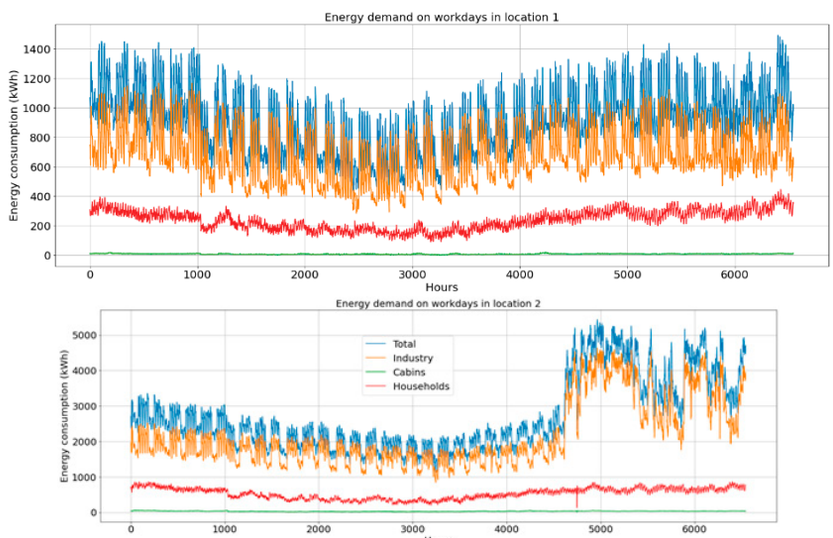


Figure 3. Real-world time series of consumption for location 1 (upper) and 2 (lower). The periods of holidays are removed, leaving a time series of 6544 samples instead of 8129 samples (1 March 2019 to 1 February 2020).

At each location, we analyze the consumption of households and industry separately. The consumption of cabins is too low in both locations to significantly affect the overall consumption patterns in the communities and the occurrence of power outages. Therefore, the consumption of cabins is disregarded in this study.

4.2. The Industry Time Series in Location 1

Statistical analysis of real-world time series is important to gain a better understanding of the consumption data and to interpret the results. Additionally, statistical models such as ARIMA assume the data to be stationary, which implies that the trend, seasonality, and short-term correlations must be removed before feeding the data into the models. On the other hand, machine-learning models do not require the time series to be stationary [64]. By removing the trend and the seasonality from the data, the neural networks can exploit all their resources to predict the "difficult" component of the time series. Such components are fast and noise-like oscillations.

As an example, in the following, we show how the trend and the main seasonality are computed for the industry sector in location 1. The trend is computed as the running mean with a window size of one week (168 h).

The running mean of the industry consumption during workdays is plotted in Figure 4; it shows a minor decrease from timestamp 0 to 3000 while increasing from 3000 to 6000 where the industry is heading towards the fishing season during the winter period. Figure 4 also shows that the consumption is cyclic with a similar load pattern every week.

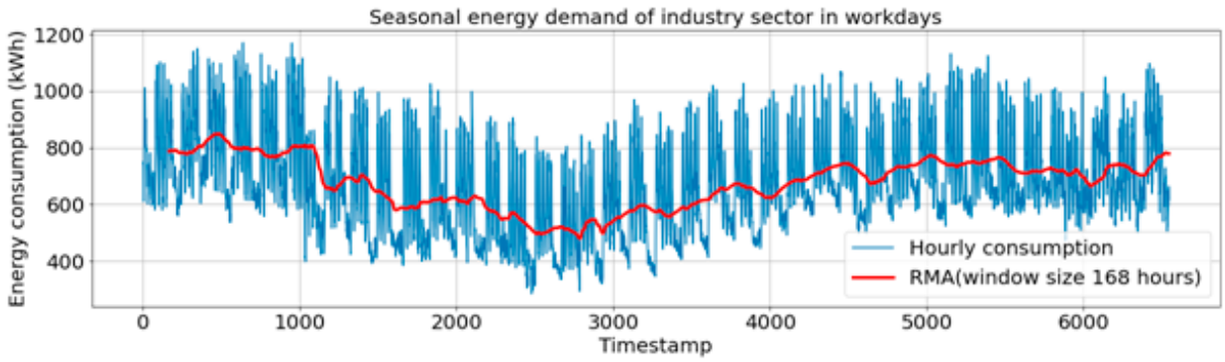


Figure 4. The industry energy consumption during workdays.

In order to identify the main consumption pattern in the time series, some statistical analyses and pre-processing are required. First, the autocorrelation function (ACF) and partial autocorrelation functions (PACF) are computed to get a better understanding of the data provided. In addition, the ACF and PACF are interpreted to investigate whether the data are stationary or not. When making predictions with statistical models such as ARIMA, the time series are required to be stationary. If the ACF and PACF plots show no correlations, the time series are stationary and ARIMA predictions can be performed [64]. The representation of the main consumption pattern together with the ACF and PACF for the industry consumption is shown in Figure 5.

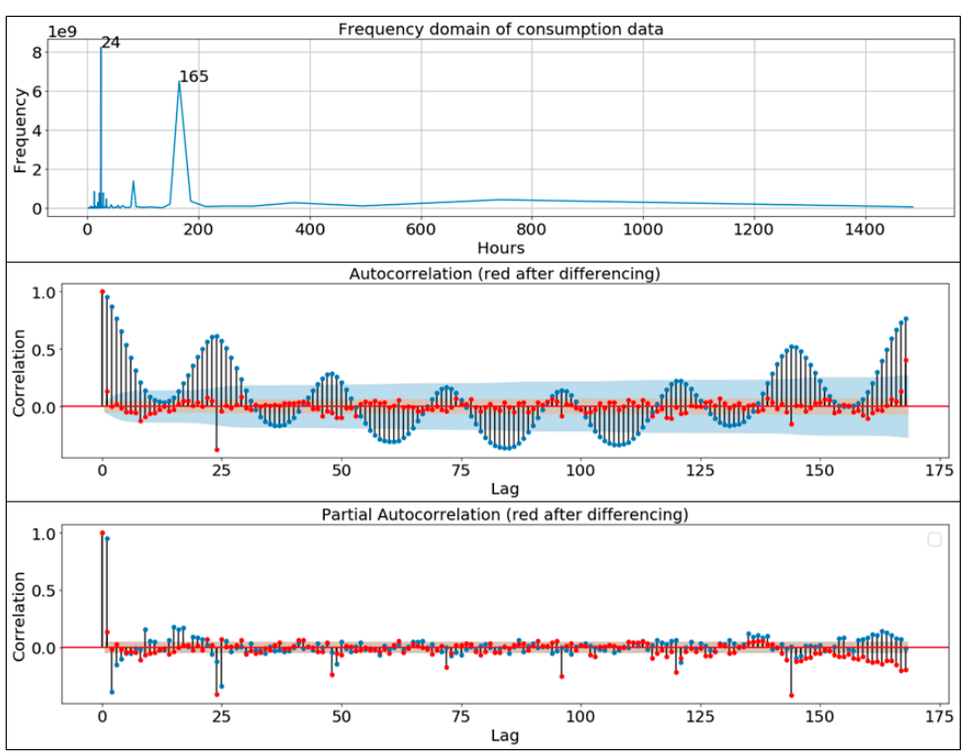


Figure 5. Frequency domain of consumption data by Fourier transformation with autocorrelation function (ACF) and partial autocorrelation functions (PACF) functions. The correlation outside the standard deviations are correlations and not a statistical fluke. The red color represents the ACF and PACF after differencing the time series.

The ACF and PACF show strong correlations outside the 95% confidence interval (depicted as a blue area in Figure 5) at 24 h and between 150 and 175 h. We note that the values outside of the blue cone are very likely actual correlations.

Another way to identify the seasonality is through the Fourier transform, which allows identifying the main periodicities in the time series [65]. The Fourier transformation shows that the main seasonality is 24 h, which is expected as we have hourly data and the consumption pattern follows a daily cycle. The time series has also a strong seasonality at 165 h, which corresponds to a week. This shows that the industry sector is strongly driven by weekly consumption patterns in addition to the daily-varying consumption.

5. Results and Discussion

5.1. Result of Short-Term Predictions (1-h Forecasting Horizon)

The resulting predictions on 1-h ahead energy demand obtained by statistical approaches and neural networks are reported in Figure 6.

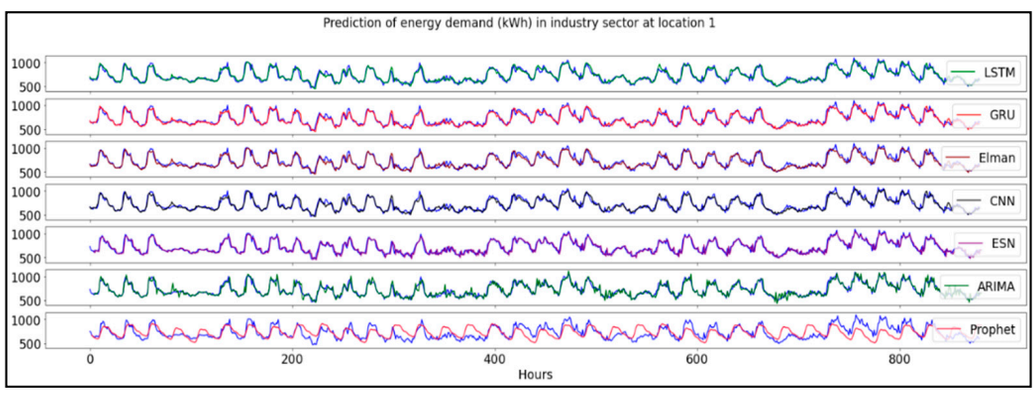


Figure 6. Short-term prediction for industry sector at location 1.

5.1.1. Industry Energy Prediction for Location 1

Figure 6 indicates that all models achieve good performance in predicting the energy consumption, except for the Prophet model. This model clearly underperforms compared to the rest of the models as it makes larger prediction errors compared to the actual consumption. This indicates that the Prophet model performs poorly in predicting the energy load profiles. For this reason, the Prophet model is not considered further in this study. The other statistical model, ARIMA, predicts the load consumption data with high accuracy (NRMSE = 0.070), but it overestimates consumption patterns. We experience larger errors from ARIMA predictions when the consumption is low.

All the neural networks used in this study achieve accurate results, and the ESN model outperforms the rest with an NRMSE result of 0.012. Among the trainable neural networks, the GRU and Elman underperform compared to the rest with an NRMSE of 0.074. All results are provided in Table 1.

We also note that all the neural networks seem to capture the peak loads with high accuracy. At peak loads, strategies for demand response by use of accurate forecasts are fundamental. The strategies are developed to plan what measures can be carried out to create an immediate change in the energy load profile. For the particular industry analyzed in this study, reducing the peak load is critical to avoid production stops resulting from power outages. The capability of capturing peak loads indicate that the neural networks are suitable tools when developing demand response strategies. Such strategies could contribute to avoiding production stops in periods where the industry operates at high loads.

Table 1. Results from 1-h prediction for all cases with all models.

Prediction Results on 1-h Forecasting Horizon				
Case 1: Train on Location 1 to Predict Location 1 and 2				
	Location 1		Location 2 (Transferability)	
	Company	Household	Company	Household
Model	NRMSE	NRMSE	NRMSE	NRMSE
LSTM	0.066	0.055	0.084	0.096
GRU	0.074	0.047	0.084	0.063
Elman	0.074	0.047	0.085	0.063
CNN	0.068	0.048	**0.075**	0.062
ESN	**0.012**	**0.013**	0.096	**0.034**
ARIMA	0.070	0.039	0.368	0.110
Prophet	0.160	0.093	-	-
Average	0.075	0.049	0.132	0.071
Case 2 (Validation): Train on Location 2 to Predict Location 2 and 1				
	Location 2		Location 1 (Transferability)	
	Company	Household	Company	Household
Model	NRMSE	NRMSE	NRMSE	NRMSE
LSTM	0.045	0.035	0.149	0.064
GRU	0.046	0.036	0.129	0.047
Elman	0.046	0.036	0.129	0.047
CNN	0.050	0.029	**0.075**	0.046
ESN	**0.032**	**0.004**	0.395	**0.040**
ARIMA	0.052	0.027	0.843	0.215
Prophet	0.220	0.076	-	-
Average	0.070	0.034	0.286	0.077

5.1.2. Short-Term Transferability Predictions of Industry Consumption at Location 2

To evaluate the transferability of the models in terms of the capability of predicting the load time series associated with different energy grids we perform the following experiments:

- the energy consumption at location 2 is predicted with the models trained on the time series on location 1;
- the energy consumption at location 1 is predicted with the models trained on the time series on location 2.

When examining the transferability by predicting the energy consumption at location 2 with the models trained on the time series on location 1, there is a larger difference in terms of prediction accuracy.

Regarding transferability predictions, the data reported in Figure 7 illustrate how the ARIMA model is no longer able to predict the load profile. In addition, ARIMA does not capture any of the actual consumption patterns. This could be because the statistics between the two-time series are significantly different.

On the other hand, all the neural networks we considered in this study seem to capture the dynamics of the load profile with high precision. Figure 7 shows that the ESN model systematically underestimates the energy consumption. This could be due to the property of the dynamical system which changes from one location to the other. When making predictions, the ESN uses a very large reservoir, which implies many parameters in the readout layer. While this helps to achieve a high prediction accuracy at location 1, the ESN can tend to over-fit the data of location 1 used for training. For this reason, the ESN makes more errors when transferred to the second location. On the other hand, the other RNNs have fewer parameters and the resulting model can generalize better to the data of the second location (see Supplementary Materials Section for more details about each model). In fact, the results show that the CNN is the model that achieves the highest prediction accuracy in transferability predictions with an NRMSE of 0.075, slightly better than LSTM with a result of NRMSE = 0.084.

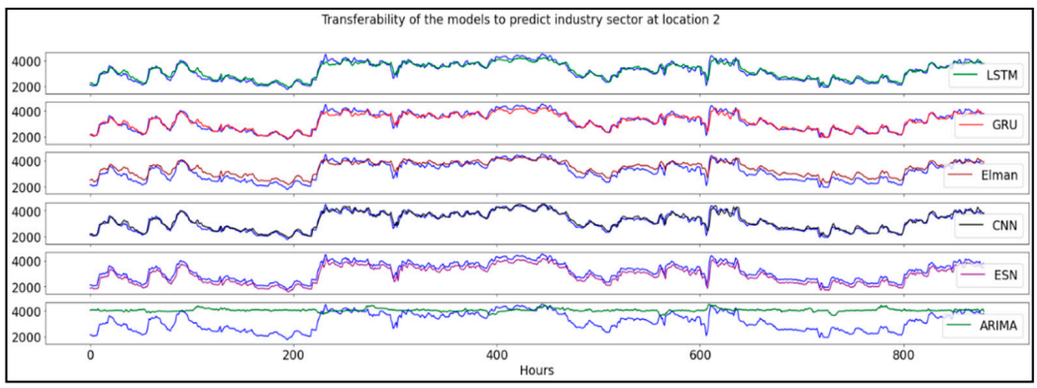


Figure 7. Short-term transferability predictions. The models are trained on the time series at location 1 to predict the energy consumption at location 2.

5.1.3. All Periods and Sectors for Both Locations

The NRMSE for both household and industry sectors and both locations are given in Table 1. The highest reported prediction accuracy is in bold. The results show that the ESN model provides the highest accuracy in predicting the energy consumption one hour onward at both locations.

Looking at the average prediction accuracy for the two sectors studied, the error is larger in the industry sector than for the household sector. The reason may be the fact that the industry sector has higher energy consumption in terms of magnitude. In addition, the larger differences between maximum and minimum load thorough the day could make the load profile more difficult to capture with high accuracy. The Fourier transformations show that the seasonality for each sector is different; the main seasonality for the household sectors was 24 and 12 h (see Supplementary Materials Section), while for industry sectors, the main seasonality were 24 and 165 h, which in turn also could affect the accuracy of the models and should be considered when performing predictions.

For transferability predictions, the ESN still outperforms all the other models in the household sector. However, when making transferability predictions in the industry sector, the CNN outperforms all other models at both locations.

Looking at the validation case where the models are trained on location 2 to predict at location 2 and location 1, the results are the same: The ESN model outperforms the other models when predicting energy demand in location 2 for both sectors. For transferability predictions, the ESN outperforms the other models in the household sector, while the CNN outperforms the other models in the industry sector.

The results show that the ESN and CNN models can predict the dynamics of the time series with high accuracy at several locations, even when trained on one source-dataset.

5.2. Predicting at Longer (2, 6, 12, 24, and 165-h) Forecasting Horizons

In some cases, it is necessary to have a prediction tool that can predict longer than one hour ahead [66]. One-hour predictions are useful when examining what demand response measures can be taken to create an immediate change in the energy load profile.

When planning for strategies in the longer term to create a more stable energy system, it is necessary to have a prediction tool that can provide accurate forecasts days and weeks in advance [66]. To investigate whether the models can predict at longer forecasting horizons, the models are tested on the industry time series for location 1, now with longer-term forecasting horizons of 2, 6, 12, 24, and 165 h. Without changing the order of the ARIMA model or any configurations in the neural networks, the results in terms of NRMSE are the following.

Again, the bold values indicate the model with the highest prediction accuracy. Table 2 shows that the ARIMA model outperforms all the other models at the 24-h forecasting horizon. The ESN model provides the highest prediction accuracy at the 1-h forecasting horizon, while the lowest accuracy is for the 12-h forecasting horizon. However, at the

24-and 165-h forecasting horizons, the accuracy increases again, which is in line with the main seasonality of 24 and 165 h observed by means of the Fourier transform discussed in Section 4.2. Since we remove the seasonality, the ARIMA model is not that sensitive to the seasonality in the data and provides significantly better prediction result at 6 h and 12 h relative to the neural networks.

Table 2. Normalized root mean squared error (NRMSE) at short- and longer-term forecasting horizon.

Multistep Prediction Results (NRMSE)						
Timestamp	**LSTM**	**GRU**	**Elman**	**CNN**	**ESN**	**ARIMA**
1 h	0.066	0.046	0.046	0.068	**0.012**	0.070
2 h	0.071	0.072	**0.067**	0.079	0.078	0.070
6 h	0.078	0.086	0.094	**0.074**	0.163	0.083
12 h	**0.077**	0.101	0.119	0.093	0.182	0.108
24 h	0.518	0.441	0.531	0.524	0.142	**0.139**
165 h	0.164	0.181	**0.083**	0.114	0.166	0.159

Looking at the neural networks, there is no clear indication that there is a model that achieves higher accuracy than the others at longer forecast horizons. It is noteworthy that all the trainable RNNs and the CNN outperform the ESN on longer forecasting horizons, which indicates that the ESN model is most suitable for being used on short-term prediction tasks. The Elman model outperforms the other models at both 2- and 165-h forecasting horizons. Again, the CNN shows impressive results, with the highest accuracy of all models at 6 h horizon, and acceptable accuracy for all horizons except for more than 24 h ahead.

The results suggest that the Elman RNN, ARIMA, and the CNN are the preferable models for longer-term energy planning purposes. However, from the results in Section 5.1, we see that the ARIMA model is difficult to transfer. This means that ARIMA must be trained on the same location where the energy consumption must be predicted and, therefore, it is necessary to possess historical energy consumption data. If no such historical data are available for training the ARIMA model, the CNN model is suggested as the optimal one.

6. Conclusions

In this paper, we studied the application of statistical models and neural networks to perform predictions of energy demand. The paper focused on predicting future loads for two communities that experience frequent power outages due to heavy loads on the existing electricity grid. Since the most important task is to predict the consumption when the load is high and there is a risk of interruptions in the power supply, we focused on energy consumption during workdays. Therefore, the periods with low consumption during holidays was removed, and we analyzed a time series of 6544-time steps.

We performed statistical analysis on both aggregated household and industry sectors and investigated the autocorrelation and partial autocorrelation functions before prediction was performed. To determine the seasonality in consumption patterns within each sector, the time series were transformed from the time domain to frequency domain by Fourier transformations. From the Fourier transforms it is clear that the industry sectors have a strong seasonality every 24 h and every 165 h, i.e., the seasonality of the industry consumption is strongly dependent on the weekly pattern, while the households have a strong seasonality every 24 h and 12 h (see Supplementary Materials Section).

The results of the predictive models in terms of NRMSE show that the ESN provides the highest accuracy when making short-term predictions. In addition, the simplicity in implementation and the fast training procedure makes the ESN model an appealing instrument for time series prediction.

To evaluate the transferability of each model, we trained each model on location 1 (or 2) and used the trained models to predict the energy consumption at location 2 (or 1).

For transferability predictions, the ESN model does also provide high accuracy results when predicting at location 2 by use of the model trained on location 1. However, when the differences in energy consumption in terms of magnitude are large, the ESN model systematically underestimates the load at location 2, which has significantly larger energy consumption. For the industry sector, the CNN model outperforms all the other models and achieves impressive results in the transferability prediction. In particular, the CNN model captured both the magnitude and consumption pattern with remarkable high accuracy. In the household sector, where differences in energy consumption in terms of magnitude were lower, the ESN model again outperforms the other models.

For longer-term predictions, the ARIMA model outperforms all models at all 24-h forecasting horizons. However, the ARIMA model is limited in that it must be trained on the same time series that are being predicted. Therefore, the ARIMA model cannot be transferred effectively. In contrast, the neural networks, and the CNN model especially, perform well when transferred to a second location. In addition, the CNN model achieves accurate predictions at multiple forecasting horizons.

The study suggests that productive future research can be undertaken to investigate the possibility of combining the ESN and CNN models for long- and short-term prediction purposes. If such a model were able to achieve high prediction results in multiple locations at both long-and short-term, it would be a valuable prediction tool to use for energy planning purposes at several locations and sectors with different time series dynamics.

Supplementary Materials: The following are available online at https://www.mdpi.com/1996-1073/14/4/798/s1, Figure S1: Energy demand for the aggregated household sector. Figure S2: ACF and PACF plots for the energy consumption in households' sector. The correlation outside the standard deviations are correlations and not a statistical fluke. Figure S3: ACF and PACF plots for the residuals after predictions with the ARIMA(1,0,1) model. Table S1: Each hyperparameter is searched in the interval [min,max]. The parameters in Table S1 are the following: Order of the autoregressive term (p), order of differentiation (d), and order of moving average term (p). The optimum hyperparameter configurations for each sector are selected as the one yielding the highest prediction accuracy on the validation set. The ARIMA(1,0,1) configuration are the model providing highest prediction accuracy for both industry and household sector. Table S2: The hyperparameter configuration for the trainable RNNs are specified as: Number of layers (L), number of units per layer (n). The hyperparameter configuration are trained over 50 epochs. Table S3: The hyperparameter configuration for the CNN are: Number of layers (L), number of units per layer (n), the convolutional kernel size (k). The dilation rate d_r, specifies how each convolutional layer L, are dilated with a factor a factor 2i. Here i are the specific layer of the network. The hyperparameter configuration are trained over 50 epochs. Table S4: Each hyperparameter is searched in the interval [min,max]. The parameters in the table are the following: Neurons in the reservoir (N_r), connectivity (R_c), noise in the state uptdate (ξ), spectral radius (ρ), the scaling of input, teaching and feedback weights (ω_i, ω_o, ω_f), and regression parameter C. The optimum hyperparameter configurations for each sector are selected as the one yielding the highest prediction accuracy on the validation set.

Author Contributions: O.F.E. performed all analyzes and had the main responsibility for writing the manuscript. F.M.B. contributed to the inception of the study and supervised in the analysis. H.A. contributed by improving the English and the structure of the manuscript. M.H. contributed with the data collections for the analyzes in the study. Y.-C.C. contributed with analyzing the data. M.C. assisted in the analysis and interpretation of the results and contributed to the inception of the manuscript O.F.E., F.M.B. and M.C. wrote the manuscript with input from all authors. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by UiT-the Arctic University of Norway, grant number 310026. The APC was funded by 310026.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: O.F.E. and M.C. acknowledge the support from the research project "Transformation to a Renewable & Smart Rural Power System Community (RENEW)", connected to the Arctic Centre for Sustainable Energy (ARC) at UiT-the Arctic University of Norway through Grant No. 310026. We thank Ishavskraft AS for providing the necessary datasets for the studies in this paper. We thank Maritsa Kissamitaki for designing Figure 1.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. De Gooijer, J.G.; Hyndman, J.R. 25 years of time series forecasting. *Int. J. Forecast.* **2006**, *22*, 443–473. [CrossRef]
2. Simchi-Levi, D.; Simchi-Levi, E.; Kaminsky, P. *Designing and Managing the Supply Chain: Concepts, Strategies and Cases*; McGraw-Hill: New York, NY, USA, 1999.
3. Bunn, D. Forecasting loads and prices in competitive power markets. *Proc. IEEE* **2000**, *88*, 163–169. [CrossRef]
4. Ruiz, P.A.; Gross, G. Short-term resource adequacy in electricity market design. *IEEE Trans. Power Syst.* **2008**, *23*, 916–926. [CrossRef]
5. Child, M.; Kemfert, C.; Bogdanov, D.; Breyer, C. Flexible electricity generation, grid exchange and storage for the transition to a 100% renewable energy system in Europe. *Renew. Energy* **2019**, *139*, 80–101. [CrossRef]
6. International Renewable Energy Agency. *Power System Flexibility for the Energy Transition, Part 2: IRENA FlexTool Methodology*; IRENA: Abu Dhabi, United Arab Emirates, 2018.
7. Alstone, P.; Gershenson, D.; Kammen, D.M. Decentralized energy systems for clean electricity access. *Nat. Clim. Chang.* **2015**, *5*, 305–314. [CrossRef]
8. Olauson, J.; Ayob, M.N.; Bergkvist, M.; Carpman, N.; Castellucci, V.; Goude, A.; Lingfors, D.; Waters, R.; Widén, J. Net load variability in Nordic countries with a highly or fully renewable power system. *Nat. Energy* **2016**, *1*, 16175. [CrossRef]
9. Bordin, C.; Thomasgard, A. SMACS MODEL, a stochastic multihorizon approach for charging sites management, operations, design, and expansion under limited capacity conditions. *J. Energy Storage* **2019**, *26*, 100824. [CrossRef]
10. Orehounig, K.; Evins, R.; Dorer, V. Integration of decentralized energy systems in neighbourhoods using the energy hub approach. *Appl. Energy* **2015**, *154*, 277–289. [CrossRef]
11. Ringkjøb, H.-K.; Haugan, P.M.; Nybø, A. Transitioning remote Arctic settlements to renewable energy systems—A modelling study of Longyearbyen, Svalbard. *Appl. Energy* **2020**, *258*, 114079. [CrossRef]
12. Olkkonen, L.; Korjonen-Kuusipuro, K.; Grönberg, I. Redefining a stakeholder relation: Finnish energy "prosumers" as co-producers. *Environ. Innov. Soc. Transit.* **2017**, *24*, 57–66. [CrossRef]
13. Liu, N.; Yu, X.; Wang, C.; Li, C.; Ma, L.; Lei, J. Energy-sharing model with price-based demand response for microgrids of peer-to-peer prosumers. *IEEE Trans. Power Syst.* **2017**, *32*, 3569–3583. [CrossRef]
14. Morstyn, T.; Farrell, N.; Darby, S.J.; McCulloch, D.M. Using peer-to-peer energy-trading platforms to incentivize prosumers to form federated power plants. *Nat. Energy* **2018**, *3*, 94–101. [CrossRef]
15. An, J.; Lee, M.; Yeom, S.; Hong, T. Determining the peer-to-peer electricity trading price and strategy for energy prosumers and consumers within a microgrid. *Appl. Energy* **2020**, *261*, 114335. [CrossRef]
16. Xiao, X.; Wang, J.; Lin, R.; Hill, D.J.; Kang, C. Large-scale aggregation of prosumers toward strategic bidding in joint energy and regulation markets. *Appl. Energy* **2020**, *271*, 115159. [CrossRef]
17. Jiang, Y.; Zhou, K.; Lu, X.; Yang, S. Electricity trading pricing among prosumers with game theory-based model in energy blockchain environment. *Appl. Energy* **2020**, *271*, 115239. [CrossRef]
18. Hafeez, G.; Alimgeer, K.S.; Khan, I. Electric load forecasting based on deep learning and optimized by heuristic algorithm in smart grid. *Appl. Energy* **2020**, *269*, 114915. [CrossRef]
19. Arcos-Aviles, D.; Pascual, J.; Guinjoan, F.; Marroyo, L.; Sanchis, P.; Marietta, M.P. Low complexity energy management strategy for grid profile smoothing of a residential grid-connected microgrid using generation and demand forecasting. *Appl. Energy* **2017**, *205*, 69–84. [CrossRef]
20. Giaouris, D.; Papadopoulos, A.I.; Patsios, C.; Walker, S.; Ziogou, C.; Taylor, P.; Voutetakis, S.; Papadopoulou, S.; Seferlis, P. A systems approach for management of microgrids considering multiple energy carriers, stochastic loads, forecasting and demand side response. *Appl. Energy* **2019**, *226*, 546–559. [CrossRef]
21. Stokland, J.; Løksa, K. Omlegging til en Framtidsrettet Nettleie, NVE. **2020**. Available online: https://www.nve.no/reguleringsmyndigheten/nytt-fra-rme/nyheter-reguleringsmyndigheten-for-energi/omlegging-til-en-framtidsrettet-nettleie/ (accessed on 9 July 2020).
22. Norges Vassdrags-Og Energidirektorat; Miljødirektoratet; ENOVA; Statens Vegvesen; Kystverket; Landbruksdirektoratet. *Klimakur 2030: Tiltak og Virkemidler mot 2030*; Miljødirektoratet: Oslo, Norway, 2020.
23. ENOVA. *Sluttrapport på Konseptutredning*; Troms Kraft Nett AS: Tromsø, Norway, 2019.
24. Herran, D.S.; Nakata, T. Design of decentralized energy systems for rural electrification in developing countries considering regional disparity. *Appl. Energy* **2012**, *91*, 130–145. [CrossRef]
25. Schäfer, M.; Kebir, N.; Neumann, K. Research needs for meeting the challenge of decentralized energy supply in developing countries. *Energy Sustain. Dev.* **2011**, *15*, 324–329. [CrossRef]

26. International Energy Agency. *SDG7: Data and Projections–Access to Electricity*; IEA: Paris, France, 2020; Available online: https://www.iea.org/reports/sdg7-data-and-projections/access-to-electricity (accessed on 15 June 2020).
27. Boute, A. Off-grid renewable energy in remote Arctic areas: An analysis of the Russian Far East. *Renew. Sustain. Energy Rev.* **2016**, *59*, 1029–1037. [CrossRef]
28. Quitoras, R.M.; Campana, P.E.; Rowley, P.; Crawford, C. Remote community integrated energy system optimization including building enclosure improvements and quantitative energy trilemma metrics. *Appl. Energy* **2020**, *267*, 115017. [CrossRef]
29. Aberilla, J.M.; Gallego-Schmid, A.; Stamford, L.; Azapagic, A. Design and environmental sustainability assessment of small-scale off-grid energy systems for remote rural communities. *Appl. Energy* **2020**, *258*, 114004. [CrossRef]
30. Statistics Norway (SSB). *Elektrisitet 10314: Nettoforbruk av Elektrisk Kraft, Etter Forbrukergruppe (GWh) (K) 2010–2019*; Statistics Norway: Oslo, Norway, 2020; Available online: https://www.ssb.no/statbank/table/10314/ (accessed on 23 October 2020).
31. Deihimi, A.; Orang, O.; Showkati, H. Short-term electric load and temperature forecasting using wavelet echo state networks with neural reconstruction. *Energy* **2013**, *57*, 382–401. [CrossRef]
32. Van Oldenborgh, G.J.; Balmaseda, M.A.; Ferranti, L.; Stockdale, T.N.; Anderson, D.L.T. Did the ECMWF seasonal forecast model outperform statistical ENSO forecast models over the last 15 years? *J. Clim.* **2005**, *18*, 3240–3249. [CrossRef]
33. Dang-Ha, T.H.; Bianchi, F.M.; Olssson, R. Local short term electricity load forecasting: Automatic approaches. In Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017; IEEE: Anchorage, AK, USA, 2017.
34. Hyndman, R.; Koehler, A.; Ord, K.; Snyder, R. *Forecasting with Exponential Smoothing: The State Space Approach*; Springer Series in Statistics; Springer: Berlin/Heidelberg, Germany, 2008.
35. Taylor, J.W. A Comparison of univariate time series methods for forecasting intraday arrivals at a call center. *Manag. Sci.* **2008**, *54*, 253–265. [CrossRef]
36. Taylor, S.J.; Letham, B. Forecasting at Scale. *Am. Stat.* **2018**, *72*, 37–45. [CrossRef]
37. Box, G.E.P.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*; John Wiley & Sons: Hoboken, NJ, USA, 2011; Volume 74.
38. Box, G.E.P.; Cox, D.R. An analysis of transformations. *J. R. Stat. Soc. Ser. B Methodol.* **1964**, *26*, 211–243. [CrossRef]
39. Alberg, D.; Last, M. Short-term load forecasting in smart meters with sliding window-based ARIMA algorithms. *Vietnam J. Comput. Sci.* **2018**, *5*, 241–249. [CrossRef]
40. Bianchi, F.M.; De Santis, E.; Rizzi, A.; Sadeghian, A. Short-term electric load forecasting using echo state networks and PCA decomposition. *IEEE Access* **2015**, *3*, 1931–1943. [CrossRef]
41. Schäfer, A.M.; Zimmermann, H.-G. Recurrent neural networks are universal approximators. *Int. J. Neural Syst.* **2007**, *17*, 253–263. [CrossRef] [PubMed]
42. Bianchi, F.M.; Maiorino, E.; Kampffmeyer, M.C.; Rizzi, A.; Jenssen, R. An Overview and Comparative Analysis of Recurrent Neural Networks for Short Term Load Forecasting. *arXiv* **2018**, arXiv:1705.04378.
43. Gasparin, A.; Lukovic, S.; Alippi, C. *Deep Learning for Time Series Forecasting: The Electric Load Case*; Cornell University,: Ithaca, NY, USA, 2019.
44. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *8*, 1735–1780. [CrossRef] [PubMed]
45. Sak, H.; Senior, A.W.; Beaufays, F. Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition. *arXiv* **2014**, arXiv:1402.1128.
46. Deihimi, A.; Showkati, H. Application of echo state networks in short-term electric load forecasting. *Energy* **2012**, *39*, 327–340. [CrossRef]
47. Varshney, S.; Verma, T. Half hourly electricity load prediction using Echo State Network. *Int. J. Sci. Res. (IJSR)* **2014**, *3*, 885–888.
48. Bianchi, F.M.; Scardapane, S.; Uncini, A.; Rizzi, A.; Sadeghian, A. Prediction of telephone calls load using Echo State Network with exogenous variables. *Neural Networks* **2015**, *71*, 204–213. [CrossRef]
49. Peng, Y.; Lei, M.; Li, J.-B.; Peng, X.-Y. A novel hybridization of echo state net-works and multiplicative seasonal ARIMA model for mobile communication traffic series forecasting. *Neural Comput. Appl.* **2014**, *24*, 883–890. [CrossRef]
50. Jaeger, H.; Haas, H. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science* **2004**, *304*, 78–80. [CrossRef]
51. Borovykh, A.; Bothe, S.; Oosterlee, C.W. Conditional time series forecasting with convolutional neural networks. *arXiv* **2018**, arXiv:1703.04691.
52. Kuo, P.-H.; Huang, C.-J. A High precision artificial neural networks model for short-term energy load forecasting. *Energies* **2018**, *11*, 213. [CrossRef]
53. Amarasinghe, K.; Marino, D.L.; Manic, M. Deep neural networks for energy load forecasting. In Proceedings of the 2017 IEEE 26th International Symposium on Industrial Electronics (ISIE), Edinburgh, UK, 19–21 June 2017; IEEE: Edinburgh, Scotland, UK, 2017; pp. 1483–1488.
54. He, W. Load forecasting via deep neural networks. *Procedia Comput. Sci.* **2017**, *122*, 308–314. [CrossRef]
55. Tian, C.; Ma, J.; Zhang, C.; Zhan, P. A deep neural network model for short-term load forecast based on long short-term memory network and convolutional neural network. *Energies* **2018**, *11*, 3493. [CrossRef]

56. Fawaz, H.I.; Forestier, G.; Weber, J.; Idoumghar, L.; Muller, P.-A. Transfer learning for time series classification. In Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018; IEEE: Seattle, WA, USA, 2018; pp. 1367–1376.
57. Xu, X.; Meng, Z. A hybrid transfer learning model for short-term electric load forecasting. *Electr. Eng.* **2020**, *102*, 1371–1381. [CrossRef]
58. Jung, S.-M.; Park, S.; Jung, S.-W.; Hwang, E. Monthly electric load forecasting using transfer learning for smart cities. *Sustainability* **2020**, *12*, 6364. [CrossRef]
59. Hooshmand, A.; Sharma, R. Energy predictive models with limited data using transfer learning. In Proceedings of the e-Energy '19: The Tenth ACM International Conference on Future Energy Systems, Phoenix, AZ, USA, 25–28 June 2019; pp. 12–16. [CrossRef]
60. Nikolay, L.; Yu, J.; Rajagopal, R. *Applied time-series transfer learning. Stanford, CA, USA. Workshop track*; ICLR; Standford university: Standford, CA, USA, 2018; pp. 1–4.
61. He, Q.-Q.; Pang, P.C.-I.; Si, Y.-W. Transfer learning for financial time series forecasting. In Proceedings of the PRICAI 2019: Trends in Artificial Intellegence, Yanuca Island, Fiji, 26–30 August 2019; Nayak, A., Sharma, A., Eds.; Lectore Notes in Computer Science. Springer: Cham, Switzerland, 2019; Volume 11671. [CrossRef]
62. Holmes, S. RMS Error, Stanford. 28 November 2000. Available online: https://statweb.stanford.edu/~{}susan/courses/s60/split/node60.html (accessed on 9 July 2020).
63. Ishavskraft AS. Om Oss. Available online: https://www.ishavskraft.no/om/ (accessed on 21 April 2020).
64. Hyndman, R.J.; Athanasopoulos, G. *Forecasting: Principles and Practice*, 2nd ed.; OTexts: Melbourne, Australia, 2018; Available online: https://otexts.com/fpp2/ (accessed on 18 January 2021).
65. Omar, K. Deconstructing Time Series Using Fourier Transform, Medium. Available online: https://medium.com/@khairulomar/deconstructing-time-series-using-fourier-transform-e52dd535a44e (accessed on 9 July 2020).
66. Hong, T.; Fan, S. Probabilistic electric load forecasting: A tutorial review. *Int. J. Forecast.* **2016**, *32*, 914–938. [CrossRef]

energies

MDPI

Article

Predicting Energy Demand in Semi-Remote Arctic Locations

Odin Foldvik Eikeland [1], Filippo Maria Bianchi [2], Harry Apostoleris [3], Morten Hansen [4], Yu-Cheng Chiou [1] and Matteo Chiesa [1,2,*]

[1] Department of Physics and Technology, UiT the Arctic University of Norway, 9037 Tromsø, Norway; odin.f.eikeland@uit.no (O.F.E.); yu.cheng.chiou@uit.no (Y.-C.C.)
[2] Department of Mathematics and Statistics and NORCE, The Norwegian Research Centre, UiT the Arctic University of Norway, 9037 Tromsø, Norway; filippo.m.bianchi@uit.no
[3] Laboratory for Energy and NanoScience (LENS), Masdar Institute Campus, Khalifa University of Science and Technology, 127788 Abu Dhabi, UAE; harry.apostoleris@gmail.com
[4] Ishavskraft Power Company, 9024 Tromsø, Norway; morten.hansen@ishavskraft.no
* Correspondence: matteo.chiesa@uit.no

Supplementary material

All the models used are described briefly in this supplementary chapter, where relevant references that describe the models more comprehensive are cited. In addition, a short discussion of the different advantages and disadvantages between the statistical- and the machine-learning approaches are discussed. Then, the case study for the household sector, which follows the same methodology as for the industry case are described. In the end, all the hyperparameter configurations used for predictions are given in chapter 5.5

Citation: Eikeland, O.F.; Bianchi, F.M.; Apostoleris, H.; Hansen, M.; Chiou, Y.-C.; Chiesa, M. Predicting Energy Demand in Semi-Remote Arctic Locations. *Energies* **2021**, *14*, 798. https://doi.org/10.3390/en14040798

Received: 29 December 2020
Accepted: 27 January 2021
Published: 3 February 2021

1. Statistical models

1.1. Autoregressive Integrated Moving Average (ARIMA)

The formalism ARIMA (p,d,q) can be used to define a large class of statistical models. The parameter p indicates the order of the auto-regressive component, d represents the initial differencing of the time series, and q the order of the moving average component. The value of the parameters must be carefully selected to achieve high prediction accuracies [1]. The equations for forecasting with ARIMA (p,q,d) are constructed as:

p: order of the autoregressive term
d: order of differentiation
q: order of the moving average term

$$\begin{aligned} y_t &= Y_t, && d = 0 \\ y_t &= Y_t - Y_{t-1}, && d = 1 \\ y_t &= Y_t - 2Y_{t-1} + Y_{t-2}, && d = 2 \end{aligned}$$

where y_t is the d^{th} difference of Y, which gives that the second difference of Y as the first difference of the first difference, i.e. the time series has performed first order differencing two times. This could be necessary if the time series are not stationary after performing differencing one time. The general forecasting equation for ARIMA predictions are:

$$\hat{y} = \mu + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} - \theta_1 e_{t-1} - \cdots \theta_q e_{t-q}, \quad (1)$$

Where θ is the moving average parameter, ϕ the slope coefficient, μ are the auto-regressive constant and e are the exponential smoothing coefficient.

In order to determine the optimal order of the ARIMA model to be used, some statistical analyses and pre-processing are required. First, the autocorrelation function (ACF) and partial autocorrelation functions (PACF) are computed to get a better understanding of the data provided, and the ACF and PACF are interpreted to investigate whether the data is stationary or not. Stationarity in time series is when statistical properties such as the mean and the variance do not change over time during the observation period. When performing predictions with ARIMA models, the time series are required to be stationary. To check for stationarity, the ACF and PACF are plotted. If the plots show no correlations, the time series are stationary and ARIMA predictions could be performed. Finally, by interpreting the ACF and PACF plots, the correct orders of the AR and MA components are identified.

1.2. Prophet

The prophet-forecasting model developed by [2], uses a decomposable time series model divided into three main components [3]. The components are: Trend (h(t)), seasonality (s(t)) and holidays (h(t)). The trend function models non-periodic changes in the data and the seasonality represents periodic changes in the data (for instance daily, weekly or yearly seasonality). The holiday effect h(t) represent potentially irregular data over one or several days. The model components are added together as:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t, \quad (2)$$

Where ϵ_t are the error term representing changes that are not captured by the model, and are assumed to be normally distributed [2]. The trend are divided into two trend models, a saturating piecewise growth model and a piecewise linear model. The piecewise logistic growth model are derived to handle trend changes in the growth rate by explicitly defining change points where the growth rate are allowed to change. If the trend shows no saturating growth, the model are selected as a piecewise constant rate of growth. The seasonality it the time series are fitted by specifying seasonality models that are periodic functions of time. To capture periodic effects, the seasonality component rely on Fourier transformation that allows identifying the main periodicities in the time series that explain the consumption pattern [4]. The final component in the prophet-forecasting model, the holidays, are incorporated by assigning a dataset D_i that represent the set of past and future dates for the each holiday i. An indicator function are added, which represent whether time t is during the holiday i, and assign each holiday the parameter κ_i, which represent the corresponding change in the forecast.

When all the components are implemented in equation (2), the prophet-forecasting model in can be fitted to predict the specific task. All derivations of the different components are provided in detail in [2].

2. Neural networks

2.1. Elman Recurrent Neural Network (ERNN)

The ERNN, also known as the *Simple* RNN, is usually considered as the most basic version of RNN [5]. The more advanced RNN architectures such as GRU and LSTM can be interpreted as an extension of ERNN. The ERNN was proposed by Jeffrey L. Elman [6], where the aim was to generalize neural networks for better handling data sequences like time-series. The effectiveness of the RNNs in handling time series comes from the ability of learning of an input sequence by means of a recurrent function [7]. The layers in an ERNN are divided into; input, hidden and output layers. The input and output layers are characterized by feedforward connections, while the hidden layer contain recurrent connections. The specific ERNN processes one element of a sequence at time. At each timestamp t, the input layer process the information at $\mathbf{x}[t] \in \mathbb{R}^{N_i}$, where N_i are the number of nodes in the input layer. The input time series $\mathbf{x}$ has a total length T. In the input

layer, each component is summed with a bias vector $\mathbf{b_i} \in \mathbb{R}^{N_h}$, where N_h is the number of nodes in the hidden layer. Each component **x[t]** is then multiplied with a weight matrix $\mathbf{W}_i^h \in \mathbb{R}^{N_i \times N_h}$. Similarly, the internal state $\mathbf{h}[t-1] \in \mathbb{R}^{N_h}$ from the recurrent time interval is summarized with a bias vector $\mathbf{b_h} \in \mathbb{R}^{N_h}$, before multiplied with the weigh matrix $\mathbf{W}_h^h \in \mathbb{R}^{N_h \times N_h}$ of the recurrent connections. Then, the transformed input and past network state are combined and processed by the neurons in the hidden layers. Finally, the output of the network at timestamp t, are:

$$\mathbf{y}[t] = g\big(\mathbf{W}_h^o(\mathbf{h}[t] + \mathbf{b_o})\big), \quad (3)$$

where the output are computed through the transformation $g(\cdot)$ on the matrix of the output weights $\mathbf{W}_h^o \in \mathbb{R}^{N_r \times N_o}$ (N_o are number of nodes in the output layer). The output weights are applied to the sum of the current state **h[t]** and the bias vector $\boldsymbol{b_o} \in \mathbb{R}^{N_o}$ [5].

2.2. Long Short-Term Memory (LSTM)

The LSTM network are essentially build in a similar way as the ERNN architecture. The main difference between LSTM and ERNN is in the composition of the inner module, where the LSTM implements a more advanced internal processing unit, a cell [5,7].

The LSTM network has the same output and input as the original ERNN. However, internally it implements a gated system that controls the neural information. The key feature of the gated networks, which makes the LSTM as a widely used neural network, is the ability to try to solve the vanishing gradient problem by not imposing any bias towards recent observations. This provides the ability that the LSTM can maintain its internal memory unaltered for long time intervals [5,7].

2.3. Gated Recurrent Units (GRU)

The GRU networks are a simplified version of LSTM. The difference between GRU and LSTM network, is that in GRU, the forget and input gates are combined and merged into a single update. This controls how much each hidden unit can remember or forget. Therefore, the GRU network ends up having two gates compared to LSTM that has three gates [5,7]. Several works shows that GRUs can perform comparably to LSTM, but generally train faster due to lighter computation [8,9].

2.4. Convolutional Neural Network (CNN)

CNNs are a class of neural networks designed to work with data that can be structured in a grid-like topology [7,10]. CNNs has been widely used for image recognition and classification, but are also suitable for forecasting univariate time-series. The CNNs are based on a discrete convolution operator, consisting of an input vector **x,** kernel **w,** and an output **f.** The convolution operator produces the output by sliding the kernel over the input vector. Each element in the output feature is obtained by summing up the result of the element-wise multiplication between the input patch and the kernel. The number of kernels used in the convolutional layer determines the depth of the output volume. In this study, the CNN are applied on a univariate energy demand time-series $\mathbf{x} \in \mathbb{R}^{n_T}$**,** with a one-dimensional kernel $\mathbf{w} \in \mathbb{R}^{\kappa}$. The output feature of the 1D CNN is:

$$f(i) = (\mathbf{x} \times \mathbf{w})(\mathbf{i}) = \sum_{j=0}^{k-1} x(i-j)\, w(j), \quad (4)$$

where i represents the i^{th} element of the convolution between **x** and **w** [7]. To better handle historical data, the original CNN are dilated to be able to learn long-term dependencies in the time-series. The dilated CNN proposed first time by the authors in [11], which named the dilated CNNs as Temporal Convolutional Network (TCN). In this study, three CNN layers are dilated by a dilation factor d, growing from d=1, d=2 and d=4. In each layer the kernel size is k=3.

The output from the TCN using the dilation factor d is:

$$f(i) = (\mathbf{x_d} \times \mathbf{w})(\mathbf{i}) = \sum_{j=0}^{k-1} x(i - dj)\, w(j), \quad (5)$$

All neural networks considered here (ERNN, LSTM, GRU and CNN) have been implemented with Python in Keras [12], with Tensorflow as backend [13].

2.5. Echo State Network (ESN)

The ESN consists of a large and untrained recurrent layer of nonlinear units in addition to a linear memory-less read-out layer. The full explanation of the underlying mechanisms for all components in the ESN architecture are explained in details in review paper by Bianchi et.al. [5].

To optimize predictions with ESN, the hyperparameter configuration must be tuned in order to maximize the prediction accuracy. Hyperparameters control the structure and the capacity of the model, determines how the network is trained, and must be specified before the ESN is trained to solve the prediction task [14]. In order for proper selection of hyperparameters to achieve accurate prediction results, the dataset analyzed are divided into three parts; training, validation and testing part [15,16].

The training set is used to fit the trainable parameters of the ESN, which are the weights of the readout. The generalization capability of the trained model, measured as the prediction accuracy obtained on future time steps of the energy load time series, is first evaluated on the validation set [15]. Since the weights of the ESN are not trained, the ESN performance is particularly sensitive to the choice of the hyperparameters.

Hyperparameters are selected to yield the highest mean accuracy on the validation set. Different strategies can be used to search for the optimal hyperparameters. Here, we tune the hyperparameters by performing a grid search over a large number of configurations (768), over 8 different hyperparameters. The hyperparameters that are optimized with cross-validation are; number of neurons in the reservoir (N_r), spectral radius (ρ), the regression parameters (C), noise in the state update (ξ), connectivity (R_c), and finally the scaling of the input, teaching and feedback weights ($\omega_i, \omega_o, \omega_f$).

When implementing and predicting the time series with ESN in this study, a modified version of the Python implementation provided by Løkse et al. is used [17].

3. Advantages and disadvantages of the statistical models and neural networks

A disadvantage of a model-based prediction approach such as the neural networks is that it is more difficult to interpret the results and understand the decision-process of the model. On the other hand, these models are easier to use for a practitioner and less domain-knowledge is required to use the models and obtain good results in terms of prediction. Statistical models are easier to interpret than the more advanced neural networks, and the user has more control on the procedure that generates the result. As a downside, a careful tuning that implies pre-analyses and a discrete amount of knowledge on the problem is required to achieve prediction results with high accuracy. This is in contrast to the neural networks, where no additional work except from careful tuning the hyperparameters is required when making predictions.

4. The household time series in location 1

This part of the supplementary chapter provides the same methodology as for the industry sector, this time for the aggregated household sector during workdays.

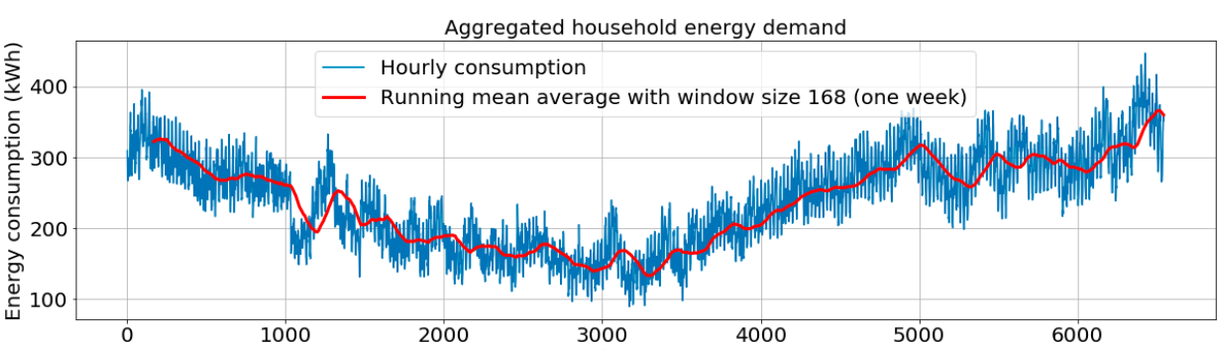


Figure S1. Energy demand for the aggregated household sector.

From Figure S1, the cyclic pattern is no longer easy to identify as for the industry sector. There is no strong weekly consumption pattern, and the consumption is much more frequent than every week as for the industry sector. During the summer period, the consumption is low as the weather is warm. The running mean average shows that the energy use evolves from high consumption early in July towards lower consumption in August. In the end of August, the consumption seems to increase again (from timestamp 3000 to 4000). The representation of main consumption pattern together with the ACF and PACF functions for the household sector are shown in Figure S2.

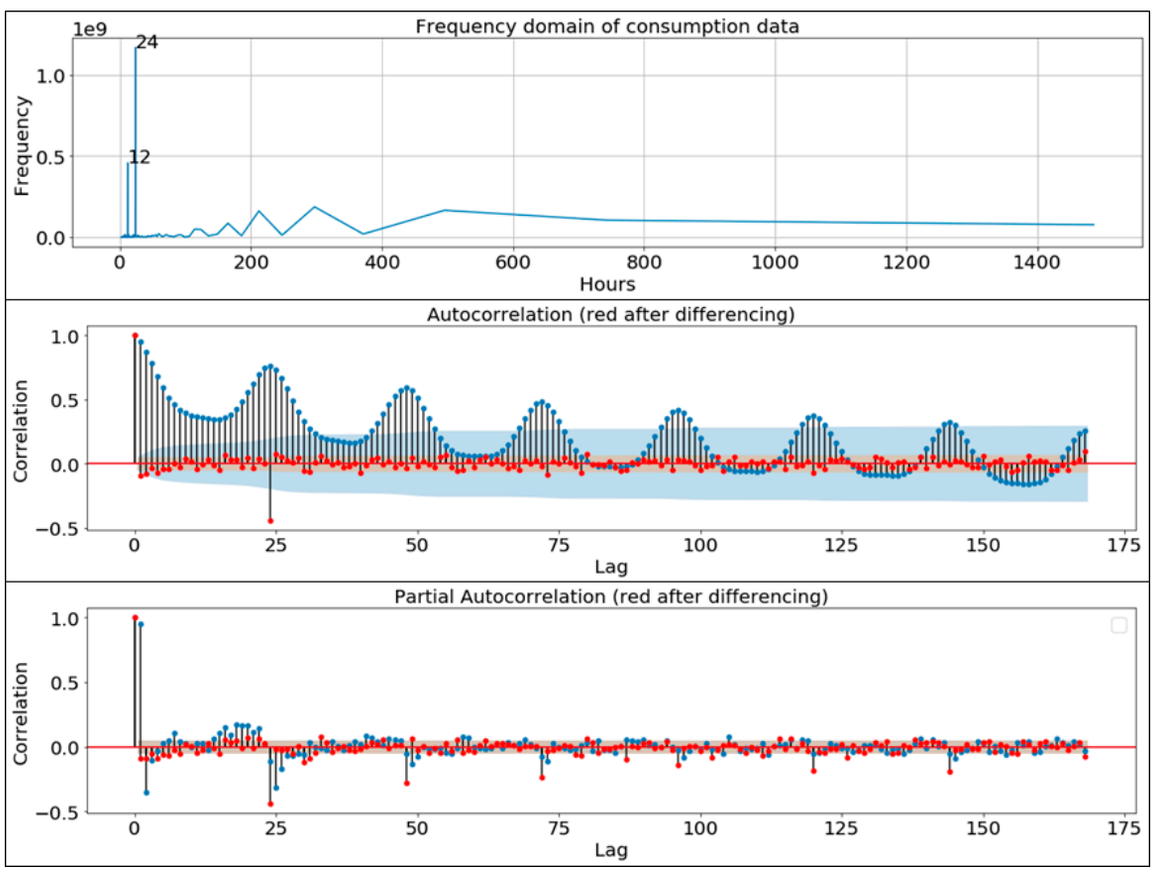


Figure S2. ACF and PACF plots for the energy consumption in households' sector. The correlation outside the standard deviations are correlations and not a statistical fluke.

The autocorrelation and partial autocorrelation show strong correlations outside the 95% confidence interval at approximately 12, 24 and 48 hours.

The Fourier transformation shows likewise a main seasonality at 24 hours. However, there is no longer a strong seasonality at 165 hours as for the industry sector. The consumption pattern is strongly driven by the daytime consumption pattern occurring every 12 hours, which in turn correlates well with the typical household daily life with routines as breakfast every morning and dinner every afternoon. The seasonality of the time series was again removed with seasonal differencing and thereafter the short-term correlation was removed by first order differencing and the ACF and PACF functions were plotted again, indicated by red colors in Figure S2.

5. Hyperparameter configurations and training of the models

For each model, the optimal hyperparameter configuration is searched on the validation set. In particular, we select as the optimal model the one yielding the highest prediction accuracy on the validation set. Then, the performance of the optimal model is evaluated on the test set.

5.1. ARIMA

To perform predictions with the ARIMA model, the time series need to be stationary. The seasonality of the time series is removed with seasonal differencing and thereafter the short-term correlation was removed by first order differencing. To check if the time series become stationary after differentiation, the ACF and PACF functions are plotted again, and they are depicted in red in Figure S2. As there are no longer strong correlations in the ACF and PACF plot (all values are almost zero in the red figure), we can assume that the time series are now stationary and ARIMA model can be applied. As the PACF function has no correlations, it is possible to use an ARIMA (1,0,1) predictive model.

In addition to selecting the ARIMA orders by interpreting the ACF and PACF plots, a grid search of different ARIMA orders was performed to identify the configuration that yields the highest prediction accuracy. The different ARIMA orders searched are provided in Table S1. Even according to this second optimization scheme, we found that the ARIMA (1,0,1) model achieved the highest prediction accuracy in terms of NRMSE.

Table S1. Each hyperparameter is searched in the interval [min,max]. The parameters in Table S1 are the following: Order of the autoregressive term (p), order of differentiation (d), and order of moving average term (p). The optimum hyperparameter configurations for each sector are selected as the one yielding the highest prediction accuracy on the validation set. The ARIMA(1,0,1) configuration are the model providing highest prediction accuracy for both industry and household sector.

ARIMA (p,d,q)	p	d	q
min	0	0	0
max	2	2	2
Optimal	1	0	1

To check for stationarity, the ACF and PACF are given for the residual errors (predicted value subtracted from the actual value). The ACF and PACF plots for the residual errors are provided in Figure S3.

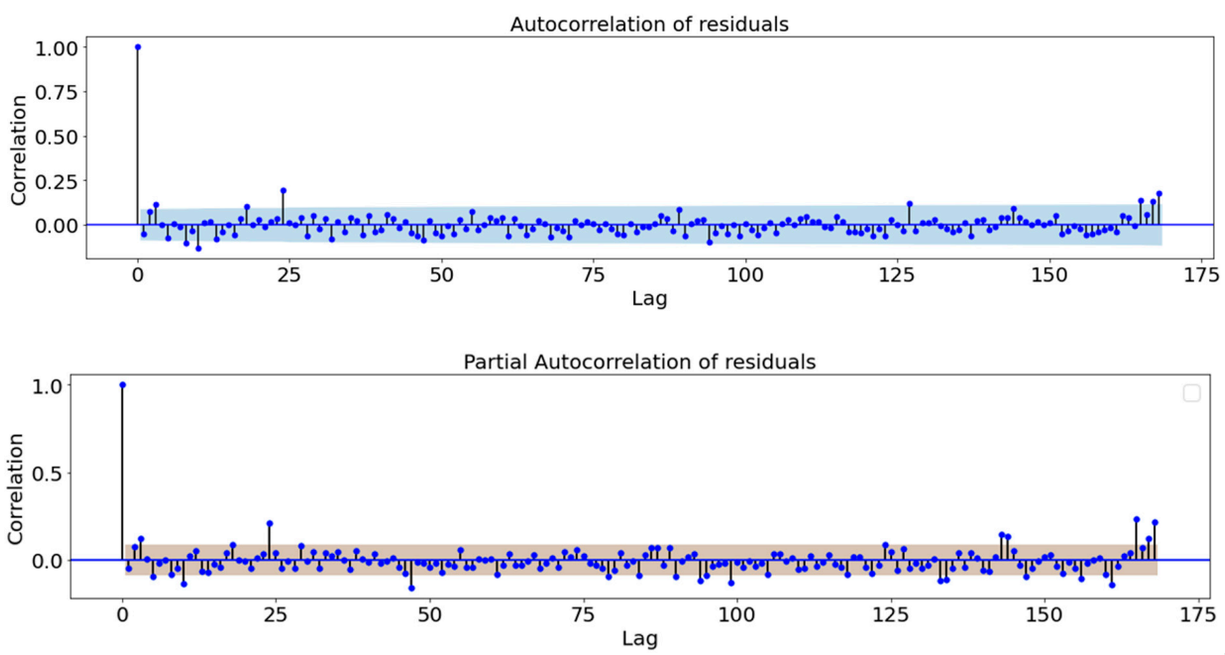


Figure S3. ACF and PACF plots for the residuals after predictions with the ARIMA(1,0,1) model.

The ACF and PACF of the residuals does not show any significant correlations. This suggests that the ARIMA (1,0,1) model is able to predict the time series.

5.2. Prophet

When making predictions with the open-source Prophet library, only the default configurations were used. The GitHub repository for making forecasts with Prophet could be found here [18]. The Prophet prediction method are developed and explained in detail in [2].

5.3. Trainable RNNs

The trainable RNNs (LSTM, GRU and Elman) are trained over several epochs on a set of configurations, where the epoch resulting in the highest accuracy on the validation set are selected for predictions on the test set. The hyperparameter configurations for the trainable RNNs are specified in Table S2.

Table S2. The hyperparameter configuration for the trainable RNNs are specified as: Number of layers (L), number of units per layer (n). The hyperparameter configuration are trained over 50 epochs.

	LSTM	**GRU**	**ERNN**
L	1	1	1
n	32	32	32
optimizer	adam	adam	adam
Learning rate	0.001	0.001	0.001
epochs	50	50	50

5.4. CNN

The CNN are trained in a similar way as for the trainable RNNs. However, inspired by the authors behind the Temporal Convolutional Network [11], we dilate the CNN to make the model able to learn long-term dependencies in the time-series. In this study, three CNN layers are dilated by a dilation factor d, growing from d=1, d=2 and d=4. In each layer, the kernel size is k=3. The hyperparameter configuration for the CNN are given in Table S3.

Table S3. The hyperparameter configuration for the CNN are: Number of layers (L), number of units per layer (n), the convolutional kernel size (k). The dilation rate d_r, specifies how each convolutional layer L, are dilated with a factor a factor 2^i. Here i are the specific layer of the network. The hyperparameter configuration are trained over 50 epochs.

	L	**n**	**k**	**d_r(L1,L2,L3)**	**optimizer**	**Learning rate**	**epochs**
CNN configurations	3	32	3	1,2,4	adam	0.001	50

5.5. ESN

When performing the predictions with ESN, the optimal hyperparameter configuration was selected with a grid search. Each hyperparameter was searched in the range specified in Table S4.

Table S4. Each hyperparameter is searched in the interval [min,max]. The parameters in the table are the following: Neurons in the reservoir (N_r), connectivity (R_c), noise in the state uptdate (ξ), spectral radius (ρ), the scaling of input, teaching and feedback weights (ω_i,ω_o,ω_f), and regression parameter C. The optimum hyperparameter configurations for each sector are selected as the one yielding the highest prediction accuracy on the validation set.

Hyperparameters	N_r	R_c	ξ	ρ	ω_i	ω_o	ω_f	C
min	300	0.15	0.0	0.5	0.1	0.25	0.0	0.001
max	500	0.45	0.01	1.0	0.4	1.0	0.1	1.0

Optimum industry location 1	300	0.15	0.0	0.9	0.1	1.0	0.0	0.001
Optimum household location 1	500	0.15	0.0	1.0	0.1	1.0	0.0	0.001

To optimize predictions with the trainable RNNs and the CNN, the optimal hyperparameter configuration could be selected by performing a grid search as for the ARIMA and ESN model. However, since we use a small number of units (32 vs the 300-500 used in the ESN model), the models are not prone to overfitting and, therefore, regularization is not required.

Additionally, since the parameters are optimized with gradient descent, the sensitivity on the hyperparameters is lower than for the ESN. Indeed, the ESN trade the precision of the gradient descent optimization with the redundancy of a large random reservoir that, inevitably, makes the model more sensitive to hyperparameters configuration.

For this reason, we can expect good performance from ERNN, LSTM, GRU, and CNN by using a fixed hyperparameters configuration.

References:

1. Nau, R. Introduction to ARIMA: Nonseasonal Models. Available online: https://people.duke.edu/~rnau/411arim.htm (accessed on 6 May 2020).
2. Taylor, S.J.; Letham, B. Forecasting at Scale. Am. Stat. 2018, 72, 37–45, doi:10.1080/00031305.2017.1380080.
3. Harvey, A.; Peters, S. Estimation procedures for structural time series models. J. Forecast. 1990, 9, 89–108
4. Omar, K. Deconstructing Time Series Using Fourier Transform, Medium. Available online: https://medium.com/@khairulomar/deconstructing-time-series-using-fourier-transform-e52dd535a44e (accessed on 9 July 2020).
5. Bianchi, F.M. Maiorino, E.; Kampffmeyer, M.C.; Rizzi, A.; Jenssen, R. An Overview and Comparative Analysis of Recurrent Neural Networks for Short Term Load Forecasting; Cornell University, Ithaca, New York: 2018, doi: arXiv: 1705.04378.
6. Elman, J.L. Finding structure in time. Cogn. Sci. 1990, 14, 179–211.
7. Gasparin, A.; Lukovic, S.; Alippi, C. Deep Learning for Time Series Forecasting: The Electric Load Case; Cornell University, Ithaca, New york: 2019
8. Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling; Cornell University: Ithaca, NY, USA, 2014. doi: arXiv: 1412.3555
9. Yin, W.; Kann, K.; Yu, M.; Schütze, H. Comparative Study of CNN and RNN for Natural Language Processing; Cornell University: Ithaca, NY, USA, 2017. doi: arXiv: 1702.01923
10. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. Proc. IEEE 1998, 86, 2278–2324.
11. Bai, S.; Kolter, J.; Koltun, V. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling; Cornell University: Ithaca, NY, USA, 2018. doi:arXiv: 1803.01271 .
12. Chollet, F.; et al. Keras, Github. Available online: https://github.com/fchollet/keras (accessed on 10 October 2020).
13. TensorFlow. An End-to-End Open Source Machine Learning Platform. Available online: https://www.tensorflow.org/ (accessed on 27 August 2020).
14. Radhakrishnan, P. What are Hyperparameters? and How to tune the Hyperparameters in a Deep Neural Network? Towards Data Science. Available online: https://towardsdatascience.com/what-are-hyperparameters-and-how-to-tune-the-hyperparameters-in-a-deep-neural-network-d0604917584a (accessed on 9 July 2020).
15. Chollet, F. Section 4.2. Evaluating Machine-Learning Models. Available online: https://livebook.manning.com/book/deep-learning-with-python/chapter-4/43 (accessed on 9 July 2020).
16. Stewart, M. Predicting Stock Prices with Echo State Networks. Towards Data Science. Available online: https://towardsdatascience.com/predicting-stock-prices-with-echo-state-networks-f910809d23d4 (accessed on 20 July 2020).
17. Løkse, S; PythonESN, GitHub. Available online: https://github.com/siloekse/PythonESN (accessed on 21 April 2020).
18. Sean, J.T and Letham, B. Forecasting at Scale—Prophet. Facebook. Available online: https://facebook.github.io/prophet/ (accessed on 15 October 2020).

/7

Paper II

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received October 29, 2021, accepted November 8, 2021, date of publication November 9, 2021, date of current version November 12, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3127042

Detecting and Interpreting Faults in Vulnerable Power Grids With Machine Learning

ODIN FOLDVIK EIKELAND[1], INGA SETSÅ HOLMSTRAND[2], SIGURD BAKKEJORD[2], MATTEO CHIESA[1], AND FILIPPO MARIA BIANCHI[3,4]

[1]Department of Physics and Technology, UiT-The Arctic University of Norway, 9019 Tromsø, Norway
[2]Arva Power Company
[3]Department of Mathematics and Statistics, UiT-The Arctic University of Norway, 9019 Tromsø, Norway
[4]NORCE The Norwegian Research Centre AS, 9294 Tromsø, Norway

Corresponding author: Filippo Maria Bianchi (filippo.m.bianchi@uit.no)

The work of Odin Foldvik Eikeland, Matteo Chiesa, and Filippo Maria Bianchi was supported in part by the Project "Transformation to a Renewable & Smart Rural Power System Community (RENEW)" under Grant 310026, and in part by the Arctic Centre for Sustainable Energy (ARC), UiT-The Arctic University of Norway.

ABSTRACT Unscheduled power disturbances cause severe consequences both for customers and grid operators. To defend against such events, it is necessary to identify the causes of interruptions in the power distribution network. In this work, we focus on the power grid of a Norwegian community in the Arctic that experiences several faults whose sources are unknown. First, we construct a data set consisting of relevant meteorological data and information about the current power quality logged by power-quality meters. Then, we adopt machine-learning techniques to predict the occurrence of faults. Experimental results show that both linear and non-linear classifiers achieve good classification performance. This indicates that the considered power quality and weather variables explain well the power disturbances. Interpreting the decision process of the classifiers provides valuable insights to understand the main causes of disturbances. Traditional features selection methods can only indicate which are the variables that, on average, mostly explain the fault occurrences in the dataset. Besides providing such a global interpretation, it is also important to identify the specific set of variables that explain each individual fault. To address this challenge, we adopt a recent technique to interpret the decision process of a deep learning model, called Integrated Gradients. The proposed approach allows gaining detailed insights on the occurrence of a specific fault, which are valuable for the distribution system operators to implement strategies to prevent and mitigate power disturbances.

INDEX TERMS Energy analytics, machine learning interpretability, power quality disturbances.

I. INTRODUCTION

Unscheduled power disturbances cause problems for customers and grid operators as they affect all customers connected to the power network, from single households to large industries [1]–[4]. Power failures might have complex and adverse socio-economic consequences in communities that are heavily reliant on the electricity supply [5], [6]. The distribution system operator (DSO) is contractually obliged to provide a reliable power supply and to compensate customers affected by power interruptions [7]. To meet the expected energy demand, the DSOs must implement management plans that account for the underlying infrastructure.

In this study, we focus on disturbances on a power grid in an Arctic region in Northern Norway, where the energy demand from local food industries has increased greatly. The growth in energy demand has resulted in more frequent power disturbances, as the current power grid is operating close to its maximum capacity. One way to improve the reliability of the power supply is to build a new distribution grid that can handle larger power demand. However, this is costly, time-consuming, has a huge environmental impact, and contradicts the vision of better utilizing the current electricity grid infrastructure* [8]. An alternative solution is to limit the failures and strengthen only the most vulnerable parts of the grid, but this requires first identifying the factors that trigger power disturbances.

The associate editor coordinating the review of this manuscript and approving it for publication was Nagesh Prabhu.

*https://www.miljodirektoratet.no/publikasjoner/2020/januar-2020/klimakur2030/

The identification of causing factors of faults in the power grid has proven to be a major challenge for the DSO [4]. However, the increased availability of energy-related data makes it possible to exploit advanced data analytics techniques to support the development of strategies for improving the reliability of the power grid [9]–[15]. Recent studies based on statistical data analysis and machine learning (ML), indicated that extreme weather conditions are often an important cause of faults in power grids [16]–[21]. However, other factors besides weather could likely affect the power quality.

In this work, we explore a wide spectrum of potential causing factors for power failures. We consider explanatory variables relative both to weather and high-resolution power-quality data. We adopt ML techniques to detect the power disturbances and to identify the factors that mostly explain the power disturbances.

This paper extends our previous study, which analyzed fault data in the Arctic power grid during the year 2020 [22]. There were important shortcomings in the data used in our previous work:

1) The machines of the local industries connected to the power grid are so sensitive to the power quality that they experience failures that are not registered in the failure-reporting system of the DSO.
2) The resolution of data in 2020 was too low (1-hour) to understand how power consumption truly affects power quality.

To address these issues, new power quality meters were installed on 19 February 2021 in the power grid under analysis. These meters log data every minute and register every small voltage variation. In addition, they provide detailed information about the power quality in the grid, such as the specific phase where the fault is registered, the magnitude of voltage variation, frequency imbalance, and the amount of flicker.

Contributions: First, we build a power faults classification dataset in collaboration with the DSO, by collecting variables that are considered as most relevant in explaining power disturbances. Then, we train different classifiers, including linear classifiers and a deep learning architecture, to detect an incoming fault from the weather and power-quality variables, registered one minute before the specific fault occurs. As shown in the experimental results, the classifiers manage to detect most of the power disturbances before their onset, demonstrating that high-resolution data from power quality meters in conjunction with weather data are highly informative.

To gain a better understanding of the relationships between the different variables and the power disturbances, we analyze the decision process of the classifiers. First, we consider traditional features selection methods, which identify which are the most important variables in the dataset that explain the fault occurrence. While such an approach gives a global overview of the variables that are, on average, the most informative in the dataset, it does not allow to reason about specific cases.

To address this challenge, we adopt a recent technique to interpret the decision process of a deep learning model, called Integrated Gradients (IG). For each individual sample, IG assigns to each feature a score, whose magnitude indicates how much the value of such feature contributes to determine the class of the sample. The proposed methodology shows that the classifiers focus on heterogeneous sets of features when processing different samples. This indicates that the occurrence of faults can be explained by multiple different patterns in the weather and power-quality variables. Our findings are valuable to the DSO for implementing strategies to prevent and mitigate power disturbances.

II. RELATED WORK AND STUDIES

There exist a vast amount of literature about the detection of different classes of power quality disturbances, such as deviation in voltage, current, and frequency signals. For example, Ref. [23] provides a comprehensive review of more than 150 research studies between 1986 and 2014 on detection and classification of power quality disturbances. In another comprehensive and more recent survey, [24] reviewed 242 papers on Power Quality and Classification (PQD&C) techniques based on digital signal processing and ML. The survey performed a comparative assessment on various PQD&C techniques by considering several criteria, such as type of data used, type of PQ disturbance, and classification accuracy.

However, fault detection and classification is a reactive process where models try to classify the fault after it has occurred. On the other hand, it is often interesting to identify the causing factors and predict the onset of a power fault. A fault prediction model should be able to quantify the likelihood of observing a fault in the next period given a set of conditions described by the explanatory variables in the model. Additionally, the identification of causing factors for faults will help the DSO to implement strategies to prevent and mitigate incoming faults.

There exist some prior relevant work on identifying causing factors for faults in the power grid. The causing factors are often divided into two different categories: i) weather conditions, and ii) other factors such as human-related activities (energy consumption).

A. WEATHER-RELATED FAULTS

Harsh and severe weather events are considered to be an important source of faults, and several studies have been conducted to address the impact of such events on power quality.

Owerko *et al.* predicted power faults in New York City by monitoring weather conditions [21]. The authors deployed a Graph Neural Network to model the spatial relationships between weather stations and improve the prediction performance.

The impact of seasonal weather on forecasting power disturbances was investigated in [25]. The authors tested the performance of the proposed models by using two different training sets: seasonal or all-year data. It was shown that,

in some cases, the prediction performance of the models improved when the training data is limited to a subset corresponding to a particular meteorological season.

The impact of weather variations and extreme weather events on the resilience of energy systems was investigated in [16]. The authors developed a stochastic-robust optimization method to consider both low impact variations and extreme events. The method was applied on 30 cities in Sweden. The results indicated that 16% drop in power supply reliability is due to extreme weather events.

Other examples of relevant work on weather-related faults can be found in Refs. [17]–[20]. In addition, several risk assessment studies on the impacts of extreme weather hazards such as earthquakes, thunderstorms, and hurricanes can be found in Refs. [26]–[32].

The works mentioned so far consider only severe weather events and disregard other factors, such as heavy energy load caused by human-related activities. Additionally, many methodologies are tested on synthetic data or on public benchmark datasets, which limits the scope of the evaluation and poses constraints on the data acquisition procedure.

B. ALTERNATIVE APPROACHES FOR FAULT DETECTION

A methodology to predict power faults by analyzing advanced measurement equipment such as Power Quality Analyzers (PQAs) and Phasor Measurement Units (PMUs.) has been proposed in [33]. The study used real-world measurements from nine PQA nodes in the Norwegian grid to predict incipient interruptions, voltage dips, and earth faults. The authors find incipient interruptions easiest to predict, while earth faults and voltage dips are more challenging to predict.

The authors in [34], compared several ML methods to predict power disturbance events such as voltage dips, ground faults, rapid voltage changes, and power interruptions. The Random Forest models achieved the highest performance and the results indicated that voltage dips and rapid voltage changes were the easiest to predict.

The challenge of detecting back-fed ground-faults has been recently addressed in [35]. The authors show that faults can be detected by integrating advanced metering infrastructure with a distribution management system. However, the proposed solution is relevant only for DSOs that adopt the OpenDSS software.

The study in [36] investigated the possibility of predicting voltage anomalies minutes in advance by using an ML model trained on historical power quality analyzers (PQA) data. The voltage data were collected from 49 measuring locations in the Norwegian power grid. The model attempted to predict voltage anomalies 10 minutes in advance based on the presence of early warning signs in the preceding 50 minutes. It was found that the time passed since the previous fault is a major factor that affects the probability of a new imminent fault.

In [37], the application of clustering and dimensionality reductions techniques to predict unscheduled events was investigated. First, the authors used several techniques to reduce the dimensionality of the data and to cluster events based on analytical features. Then, the fault events were separated from the normal operating conditions. The findings show promising results when using balanced datasets, while the predictive capability is significantly reduced in unbalanced datasets that, however, often appear in real-world case studies.

Other relevant work on fault detection based on ML techniques can be found in Refs. [38]–[43]. In addition, there is some relevant work that adopts novel ML techniques for detecting and localizing faults in the power distribution network [9]–[12].

This section presented several relevant works in predicting faults by assessing either weather effects or human activities. One of the goals of our work is to consider, at the same time, a larger amount of weather variables and electricity-related measures as potential causes of power disturbances. A close collaboration with the local DSO has provided us with valuable insights about the relevant variables that should be monitored to construct a new classification dataset. More importantly, none of the previous work has focused on interpreting the decision process of the classifier, which is key to understanding the causes of faults and can provide valuable information to improve the power grid reliability.

III. POWER FAULTS DATASET

In this study, we focus on a power grid with a radial structure located in the Arctic. A detailed description of the grid configuration is deferred to Sect. A in the Appendix. The grid is subject to frequent power faults, which could be caused by weather factors or by the strain of the infrastructure from a local industry, which dominates the load consumption in the power grid.

We prepared a classification dataset where each sample refers to a period when the grid is operating in normal conditions or to a period preceding a fault, respectively. Each sample is associated with a feature vector $x \in \mathbb{R}^{12}$ and a label $y \in \{0, 1\}$, indicating the normal condition or the imminent fault, respectively. The feature vector consists of 6 different energy-related variables and 6 different weather variables, summarized in Tab. 1. A fault is registered when there is at least a 10% drop in voltage magnitude. Further details about faults measurement, what the weather and power variables represent, and how they are collected, are described in Sect. B in the Appendix.

The dataset contains 90 samples representing reported faults ($y = 1$), which occurred in the period between 19.02.2021 to 30.04.2021. Naturally, the amount of samples associated with normal operating conditions is much larger. In addition, in normal operating conditions the values x from neighboring hours are very similar to each other. To limit the amount of class imbalance in the dataset and the redundancy in the over-represented class, we arbitrarily subsampled the non-fault class ($y = 0$) by taking 1 sample every 60. In the final dataset, there are 90 samples representing a reported

TABLE 1. Variables analyzed to detect faults in the power grid.

Feature ID	Weather variables
1	Wind speed of gust
2	Wind direction
3	Temperature
4	Pressure
5	Humidity
6	Precipitation
	Power variables
7	Difference in Frequency
8	Difference in Voltage imbalance
9	Difference in Active Power
10	Minimum Power Factor
11	Difference in Reactive Power
12	Flicker

fault and 1, 647 samples representing normal operating conditions without any power disturbance.

IV. METHODOLOGY

Our approach consists of two steps. First, we train a classifier to predict the onset of power faults given the value of the electricity and weather variables. If we obtain a high classification accuracy, we can conclude that there are strong relationships between the variables, x, and the occurrence of faults, y. Then, we use two different techniques to highlight the most informative features identified by the classifiers to solve the task.

In Sect. IV-A and IV-B, we describe which classifiers are considered in this study. In Sect. IV-C, we present an approach for interpreting the decision process of a neural network classifier.

A. LINEAR CLASSIFIERS

We consider three different linear classifiers. The first is a Ridge regression classifier, which first converts the target values into $\{1, 1\}$ and then treats the problem as a regression task [44]. The second model is Logistic regression, which uses a logistic function to approximate the probability of binary classification variable [44]. The third model is the Linear Support Vector Classification model (LinearSVC), which is a type of Support Vector Machine (SVM) [45] endowed with a linear kernel.

Due to the strong class imbalance, we configure each model to assign a class weight that is inversely proportional to the number of samples in each class. In this way, errors on the underrepresented class (faults, $y = 1$) are penalized much more than errors on the larger class (nominal condition, $y = 0$).

One advantage of using linear classifiers is that they construct a decision boundary directly in the input space, which allows to easily interpreting the decision process of the classifier. In particular, the linear models assign a weight w_i to each feature x_i in the input space: the higher w_i, the more the values of x_i impact the classification outcome. Therefore, looking at the magnitude of the weights w_i is a simple strategy to estimate the average importance of the features in the dataset for the classification task.

B. NON-LINEAR CLASSIFIERS

We consider two non-linear classifiers. The first is a non-linear SVC classifier equipped with a radial basis function kernel (RBFSVC). As for the linear models, also in this case we used class weights inversely proportional to the class frequency.

The second non-linear classifier considered is a Multi-Layer Perceptron (MLP) [46]. The MLP consists of an input layer that takes input vectors $x \in \mathbb{R}^{12}$, L hidden blocks, an output layer that generates a 2-dimensional output $o \in \mathbb{R}^2$, and a softmax activation that gives the vector of class probabilities y. Each block l consists of a dense layer with n_l units, a Batch Normalization layer [47], a non-linear activation function, and a Dropout layer [48] with dropout probability p. All trainable weights in the MLP, except the biases, are regularized with L_2-norm penalty with strength λ. Fig. 1 depicts the MLP architecture.

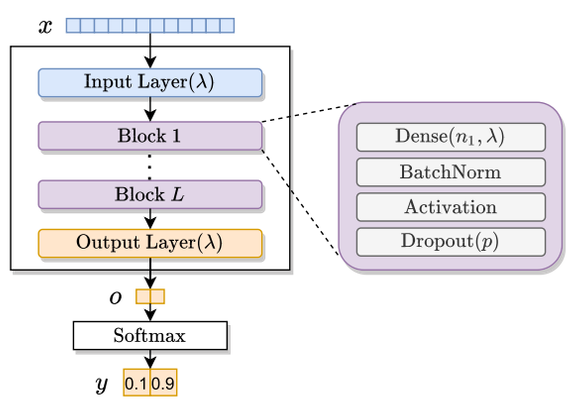


FIGURE 1. Architecture of the MLP.

The MLP is trained by minimizing a cross-entropy loss, using batches of size b, and the Adam optimization algorithm [49] with initial learning rate r. Due to the strong class imbalance in the dataset, we initially trained the MLP by weighting the loss of each sample with a value inversely proportional to the class frequency, as we did for the other classifiers. However, we found out that the MLP achieved better performance by re-sampling the minority class during training. This allows achieving class balance at the expense of introducing redundancy, by re-proposing the same samples multiple times. We also tried to achieve class balance by subsampling the majority class but, due to the small number of samples in the fault class, the total number of inputs in each training epoch was too small and the samples from the majority class were shown too few times during training.

C. INTERPRETATION OF THE MLP RESULTS WITH INTEGRATED GRADIENTS

In the following, we introduce the technique adopted to interpret the decision process of the MLP. A short review of important approaches for interpretability in deep learning,

which have been proposed over the past few years (and briefly mentioned hereafter), is deferred to Sect. C in the Appendix.

Integrated Gradients (IG) [50] is a technique proposed to satisfy two axioms, which are not jointly enforced by other existing attribution schemes (see Sect. C for details). According to the first axiom, *sensitivity*, if the input and an *uninformative baseline* differ in exactly one feature, such a feature should be given non-zero attribution. While interpretability approaches such as LRP [51] and DeepLiFT [52] ensure sensitivity due to the conservation of total relevance, gradient-based methods [53]–[56] do not guarantee the sensitivity axiom because the saturation at ReLU or MaxPool makes the score function locally "flat" with respect to some input features.

The second axiom, *implementation invariance*, states that when two models are functionally equivalent, they must have identical attributions to input features. While implementation invariance is mathematically guaranteed by vanilla gradient approaches, the coarse approximation to gradients in LRP and DeepLiFT might break this assumption.

The attribution to feature i given by IG is

$$\mathrm{IG}_i(x) ::= (x_i - x_i') \times \int_{\alpha=0}^{1} \frac{\partial F\left(x' + \alpha \times (x - x')\right)}{\partial x_i} d\alpha, \quad (1)$$

where i is an input feature, x is a sample in the dataset, x' is the uninformative baseline, and α is an interpolation constant used to perturb the features of the input sample. The above definition ensures both the desirable assumptions:

- By the Fundamental Theorem of Calculus, IGs sum up to the difference in feature scores and, thus, follow sensitivity;
- Since the IG attribution is completely defined in terms of gradients, it ensures implementation invariance.

IG has become a popular interpretability technique due to its broad applicability to any differentiable neural network model, ease of implementation, theoretical justifications, and computational efficiency.

Implementation: IG is a post-hoc explanatory technique that works with any differentiable model, $F(\cdot)$, regardless of its implementation. In this paper, we let $F(\cdot)$ be the MLP model described in Section IV-B that takes as input tensor the feature vector $x \in \mathbb{R}^{12}$ and generates an output prediction tensor, $o = F(x)$, called *logit*. In our case, $o \in \mathbb{R}^2$ and softmax(o) gives the probability of x being "fault" and "non-fault".

The baseline x' in (1) is an uninformative input used as a starting point to compute the IG attributions. The baseline is essential to interpret the IG attributions as a function of individual input features. It is important to choose a baseline that encodes as much as possible the lack of information about the target class c. In a classification task with multiple classes, we want softmax$[F(x')]_c \approx 0$. In a binary classification task, like in our case, we can chose a baseline that gives equal probability of belonging to both classes, i.e., softmax$[F(x')]_0 \approx$ softmax$[F(x')]_1 \approx 0.5$. In computer vision tasks, a black image (all pixels at 0) is commonly used as a baseline. However, in our dataset, the value 0 might actually be informative because the absence of some specific features can increase the probability of belonging to a specific class (e.g., in the absence of wind it is less likely to observe a fault). Fig. 2(left) shows that the MLP assigns with high confidence the zero-baseline x_z' to class 0 (non-fault). Therefore, different alternatives should be considered as the baseline. One option is to cast the binary classification problem into a 3-classes problem and re-train the two that assign a vector of zeros to a third, dummy class. In this way, when using the zero-baseline x_z', we would get softmax$[F(x_z')]_0 \approx$ softmax$[F(x_z')]_1 \approx 0$. Other alternatives are to use a mean-baseline, x_m', which is a vector computed as a weighted average of the features across the two classes or to use, or a random baseline x_r'. In the latter case, the final result is given by averaging the IG attributions computed from several random baselines. As shown in Figure 2, the mean baseline gives almost the same probability to classes 0 and 1, while the random baseline has the tendency to assign a strong probability to one of the two classes. Therefore, we used the mean baseline in all our experiments.

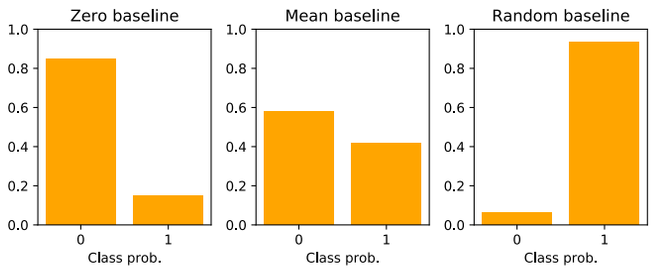


FIGURE 2. Class probabilities for different baselines on the power-faults dataset.

The default path used by the integral in (1) is a straight line in the feature space from baseline to the actual input. Since the choice of path is inconsequential with respect to the above axioms, we use the straight-line path that has the desirable property of being symmetric with respect to both x and x'. The numerical computation of a definite integral is often not tractable and is necessary to resort to numerical approximations. The Riemann trapezoidal sums offer a good trade-off between accuracy and convergence and changes (1) into:

$$\mathrm{IG}_i^{\mathrm{approx}}(x) ::= (x_i - x_i') \times \sum_{k=1}^{m} \frac{\partial F\left(x' + \frac{k}{m} \times (x - x')\right)}{\partial x_i} \times \frac{1}{m}, \quad (2)$$

where m is the number of finite steps used to approximate the integral and $\alpha \approx k/m$. The m samples $\mathcal{X} = \{x' + \frac{k}{m} \times (x - x')\}_{k=1}^{m}$ represent the linear interpolation between the baseline and the input. Fig. 3 depicts such an interpolation path from the mean-baseline to a specific sample of class "fault" in our dataset.

After generating the set of interpolated samples $\mathcal{X}$, we can compute the gradients $\frac{\partial F(\mathcal{X})}{\partial x_i}$ that quantify the relationship

between the changes in the input features and the changes in the predictions of the MLP F. Important features will have gradients with steep local slopes with respect to the probability predicted by the model for the target class. Interestingly, the largest gradient magnitudes generally occur during the first interpolation steps. This happens because the neural network can saturate, meaning that the magnitude of the local feature gradients can become extremely small and go toward zero resulting in important features having a small gradient. Saturation can result in discontinuous feature importances and important features can be missed. This is the key motivation why rather than simply using the gradients of the actual input, $\frac{\partial F(\mathcal{X})}{\partial x_i}$, IG sums all the gradients accumulated during the whole interpolation path. This concept is exemplified in Fig. 4(left), showing that the model prediction quickly converges to the correct class in the beginning and then flattens out. There could still be less relevant features that the model relies on for correct prediction that differs from the baseline, but the magnitudes of those feature gradients become really small, as shown in Fig. 4(right). The Figure is obtained using the same data of Fig. 3.

V. EXPERIMENTAL EVALUATION

After introducing the experimental setting, in Sect. V-A we compare the classification performance of the different classifiers on our dataset. Then, in Sect. V-B we first analyze the specific samples of class "fault" that are missed by the classifiers and, then, we consider two techniques to interpret the decision process of the classifiers.

Model Selection and Performance Evaluation: The linear and the SVM classifiers are implemented with the scikit-learn library,[†] while the MLP is implemented in Tensorflow.[‡] To evaluate the model performance we first shuffle the data and then perform a stratified k-fold, with $k = 5$. In each fold, 80% of the data are used as a training set, and the remaining 20% is used as a test set. The training is further divided into two parts: 80% is used to fit the model's coefficients and 20% is used as a validation set to find the optimal hyperparameters.

The hyperparameters of the linear models and the SVM are optimized with a grid search. In particular, we optimize the regularization strength in the Ridge regression classifier, Logistic regression, and LinearSVC. For the non-linear SVM classifier, we also optimize the width of the radial basis function.

For the MLP, due to the higher amount of hyperparameters and the longer training time, we used the Bayesian optimization strategy implemented in Keras Tuner[§] and evaluated a total of 5,000 configurations. In particular, we optimized the number of layers L, the number of units n_l in each layer, the L_2 regularization coefficient λ, the dropout probability p, the learning rate r, and the type of activation function (ReLU, tanh, or ELU). We used a fixed batch size $b = 32$, an early stopping with patience of 30 epochs, and we reduced the initial learning rate by a factor of $1/2$ when the validation loss was not decreasing for 10 epochs.

Before training the models, the input values x are normalized feature-wise by subtracting the mean and dividing by the standard deviation computed on the training set. The overall performance of each classification model is the average performance obtained on each test set of the 5 folds.

Performance Measures: The classification performance is measured by looking at the confusion matrix, which reports the following quantities: True Negatives (TN) – correctly identified non-faults, False Positives (FN) – non-faults predicted as faults, False Negatives (FN) – faults missed, and True Positives (TP) – faults correctly identified. To quantify the performance with a single value we use the F1 score:

$$F1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} = \frac{\text{TP}}{\text{TP} + \frac{\text{FP+FN}}{2}}. \tag{3}$$

Due to the strong class imbalance in the dataset, we compute a weighted F1 score, i.e., we weight the F1-score obtained for each class by the number of samples in that class and then we compute the average:

$$F1_{weighted} = \frac{(n_{\text{faults}} \cdot F1_{\text{faults}}) + (n_{\text{non-faults}} \cdot F1_{\text{non-faults}})}{n_{\text{faults}} + n_{\text{non-faults}}}, \tag{4}$$

where $n\_$ and $F1\_$ indicate the number of samples and classification scores for each class, respectively.

Selecting the Number of Interpolation Steps in IG: The result of the IG attribution depends on the number of steps m (see Eq. 2). One of the properties of IG is completeness, meaning that feature attributions encompass the entire prediction of the model. As a consequence, the importance score should capture the individual contribution of each feature to the prediction. Therefore, by adding together all the importance scores is it possible to recover the entire prediction value for a given sample x. In particular, we have that the variation in classification score (e.g., the probability of being a fault) is

$$\delta = \sum_i \text{IG}_i(x) - \left(F(x)_c - F(x')_c\right)$$

where $F(x)_c$ and $F(x')_c$ are the prediction scores for class c when the model takes as input x and x', respectively. Since we want the $\sum_i \text{IG}_i(x)$ to explain the whole difference in the class attributions, the number of integration steps m should be increased until when δ becomes as close as possible to zero. Following this principle, we found $m = 100$ to be sufficiently large for our experiments as it gives $\delta < 1e-2$.

A. CLASSIFICATION PERFORMANCE OF THE DIFFERENT METHODS

Here, we compare the classification performance obtained by the linear methods, SVM classifier, and the MLP. The classification performance of each model is reported in Tab. 2 in terms of average Weighted F1 score and the average number

[†] https://scikit-learn.org/
[‡] https://www.tensorflow.org/
[§] https://keras.io/keras_tuner/

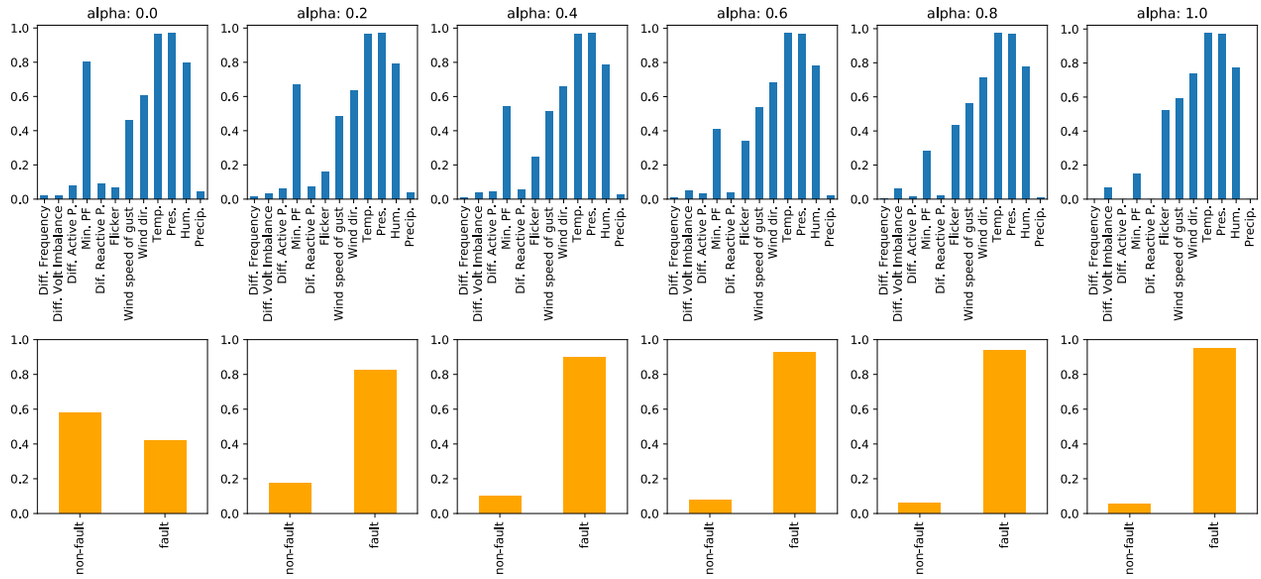


FIGURE 3. Top row: linear interpolation from the mean-baseline (left) to an actual sample of class fault (right). Bottom row: classification probabilities assigned by the MLP at each step of the interpolation.

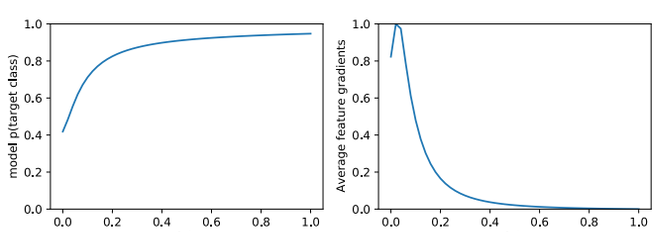


FIGURE 4. Left: target class predicted probability over α. Right: average feature gradients (normalized) over α.

TABLE 2. Classification score for different models.

	Classifier	TN	FP	FN	TP	Weighted F1 score
Linear	Ridge Classifier	272	57	4	14	0.785
	Logistic regression	275	54	5	13	0.756
	LinearSVC	276	54	5	13	0.757
Non-linear	RBFSVC	283	46	5	13	0.771
	MLP	285	45	4	14	0.803

of TN, FP, FN, and TP obtained across the 5 folds. Note that the TN, FP, FN, and TP are rounded to the closest integer.

The MLP classifier achieves top performance with a weighted F1 score of 0.803, followed by the Ridge Classifier and the SVC with RBF kernel that achieves weighted F1 scores 0.785 and 0.771, respectively. In our case study, is important to miss as few faults as possible, meaning that solutions with very few FN (missed detection) are acceptable even if the number of FP (false alarms) is significant. The MLP and Ridge Classifier provides the most promising result with 4 FN and 14 TP on average.

Finally, it is interesting to notice that linear and non-linear models achieve similar performance. This suggests that the two classes are almost linearly separable, i.e., most of the data samples can be separated reasonably well by an hyperplane in the input features space. On the other hand, the data samples that are misclassified are very entangled, and is difficult to find a decision boundary, even if is non-linear, that can correctly separate them. The good performance of the classification models motivates the feature interpretation procedure discussed in the next section.

B. ANALYSIS AND INTERPRETATION OF THE RESULTS

For the next analysis, we generate a fixed random train/validation/test split and used the same fold for each model. This allows us to analyze in detail the solution obtained by the different methods on a single test set, which contains 18 faults and 330 non-faults. Interestingly, all models fail to correctly classify as faults the same 5 data samples. A closer manual investigation on such 5 samples shows the following:

1) **2021-02-22 at 19:29:00:** is an empty measurement,
2) **2021-02-22 at 21:55:00:** is a phase-to-ground fault;
3) **2021-02-22 at 22:12:00:** is a phase-to-ground fault;
4) **2021-02-26 at 11:58:00:** is an actual fault that was missed by the classifiers;
5) **2021-03-02 at 09:29:00:** is a fault with an unusual long duration.

The first FN could have been caused by some type of error, such as a calibration error, in the measurement instruments.

In the case of a ground fault, the electrical transformers connected to the grid break, and the power that flows through the transformer flows to the ground. When the end of the electrical transformer station that contacts the ground level is on the downstream side, a ground fault occurs [35]. The ground fault is detected as a reduction of only one of the three-phase voltages. Fig. 5 depicts the phase voltages when the first ground fault occurred: it is possible to see that

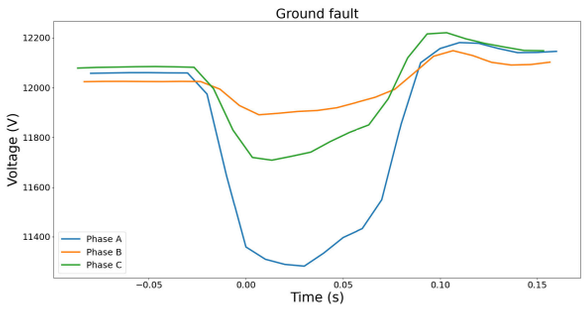


FIGURE 5. The phases in a phase-to-ground fault incident. The ground fault occurs on Phase A which is decreasing significantly (voltage drop), while the other two in the three-phase system have a minor decrease from nominal voltage level.

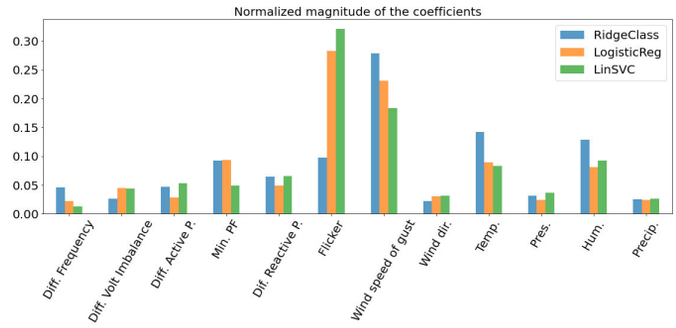


FIGURE 6. Coefficients' magnitude assigned to each feature by different linear models. High magnitude indicates that the corresponding feature is important.

Phase A decreases significantly, while the other two stay above the nominal voltage value. It is difficult to explicitly detect ground faults from only weather and electricity load measures considered as input variables, and therefore it is reasonable that the models miss the faults number 2 and 3.

Similar to the ground faults, the 4th FN could be caused by a factor not described in the weather and electricity variables. For example, it could have been caused by vegetation or animals interacting with the power lines.

Finally, the 5th FN is a fault that lasts for 200 seconds, while the usual duration of the faults is approximately 25-30 secs. This suggests that the fault is an anomaly that is not well represented in the dataset and, therefore, is difficult to be classified accurately.

To identify the most important variables that explain the faults, we try to interpret the decision process of the different models. First, we analyze the coefficients of the linear models, which give a ''global'' interpretation of the variables importance. Then, we use the IG technique for a ''local'' interpretability of the features that explain the class of a specific data sample.

1) GLOBAL INTERPRETABILITY

As discussed in Sect. IV-A, when using linear models we can interpret the magnitude of the weights assigned to the input features as the global importance of the features for the classification problem. Fig. 6 reports the feature weights learned by the three different classifiers. We observe that in each model the *Wind speed of gust* variable is always associated with a weight with a large magnitude. The Linear SVC and the Logistic Regression classifiers attribute large importance also to the *Flicker* variable, while the Ridge Regression classifiers weight the other features more uniformly and assign weights to *Temperature* and *Humidity* that are slightly larger than the weight assigned to *Flicker*.

This analysis suggests that both the industrial activity and the weather effects are important in discriminating between the fault and non-fault classes. According to the linear models, the most important among the power-related variables seems to be *Flicker*, while the *Wind speed of gust* is consistently the most explanatory weather-related variable. These observations are aligned with the experiences of the DSO and the local customers, as more faults seem to occur when there is high activity at the industries and the machines operates at full load. In addition, it has been noted that faults are more likely to occur when there is a strong wind, which could create collisions in the cables of the power lines.

2) LOCAL INTERPRETABILITY

The faults correctly classified by the different models are reported in Tab. 3, together with the confidence score of the MLP classifier. The confidence score can be interpreted as the probability that the MLP believes a sample is a fault. The MLP correctly classifies with high confidence most fault samples and assigns a probability greater than 90% to 5 out of 13 samples. As a side note, the faults do not appear to be clustered around specific days or periods, but they seem to be uniformly distributed over time.

We focus on the samples 52, 140, 227, 304, and 316 in Tab. 3, which are those classified with the highest confidence, and we use IG to identify which are the variables that are most important for the MLP to determine the correct fault class. The results are reported in Fig. 7. The top-left plot depicts the uninformative baseline, which corresponds to what an ''average'' sample in the dataset looks like. The blue bar plots represent the value of the 12 features in the 5 selected samples. Finally, the green and red bar plots are the output of the IG procedure.

The green bars indicate that a feature is important for the classification result. The higher the i-th green bar, the more the feature value x_i in the sample (blue bar) explains the classification result, compared to the value x_i' in the baseline (black bar). For example, in Sample 227, the value of Flicker is much greater than in the baseline. IG assigns a high score (tall green bar) to this difference, meaning that the MLP found important the *increment* in Flicker compared to the baseline level for deciding that Sample 227 is a fault. Similarly, the MLP found important the *decrement* in Minimum Power Factor compared to the baseline level, to classify Sample 227 as a fault.

A red bar, instead, indicates that a value x_i decreases the confidence in the classifier that the sample is actually a fault,

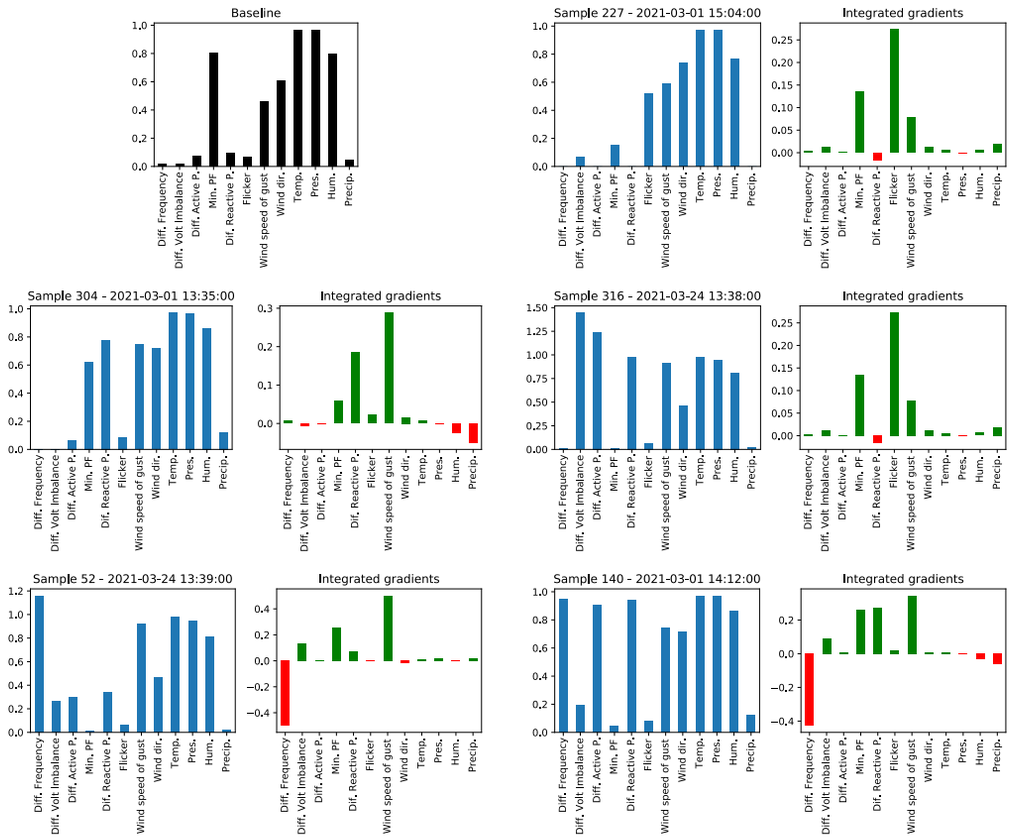


FIGURE 7. The green bars denote that a feature is important for the classification result. The higher the green bar, the more the feature value in the sample (blue bar) explains the classification result, compared to the value in the baseline (black bar). The red bars means that the value of the features in the sample decrease the confidence of the classifier that the sample is actually a fault.

TABLE 3. True positives and confidence score assigned by the MLP classifier.

	Fault ID	Confidence (MLP)	Date	Time
0	31	0.71	2021-03-02	14:30:00
1	35	0.72	2021-02-28	20:40:00
2	**52**	**0.91**	2021-03-24	13:39:00
3	**140**	**0.92**	2021-03-01	14:12:00
4	163	0.60	2021-04-07	05:16:00
5	189	0.71	2021-03-02	14:31:00
6	**227**	**0.95**	2021-03-01	15:04:00
7	235	0.88	2021-02-28	14:49:00
8	269	0.77	2021-02-28	19:28:00
9	271	0.86	2021-03-24	16:01:00
10	291	0.86	2021-03-01	14:03:00
11	**304**	**0.92**	2021-03-01	13:35:00
12	**316**	**0.96**	2021-03-24	13:38:00

compared to having a baseline value x_i'. For example, the MLP would have been even more confident that Sample 52 and Sample 140 are faults if their Difference in Frequency values would have been as in the baseline. In other words, for these two samples, the increment of Difference in Frequency is something that decreases the confidence of the classifier that they are faults.

This analysis shows that each sample has different features that are found important by the MLP for the classification. For example, Sample 227 is classified as a fault mainly because of the above-average value in Flicker; Sample 52 is a fault due to the high value of Wind speed of gust and low value in Minimum Power Factor; for Sample 304 is important that the Difference in Reactive Power is higher than average.

The Minimum Power Factor and Reactive Power are important variables that contribute to explaining the current power quality in a power grid. The Power Factor is the ratio of the working power over the apparent power and quantifies the energy efficiency: the lower the power factor, the less efficient is the power usage of the end-customer. The Reactive Power is the amount of power dissipated in the system. A high amount of reactive power in the system could affect the power quality negatively as there will be less amount of available active power that can be used by the end-customer [57]. Therefore, it is reasonable to observe a relationship between the low value in the Minimum Power factor, and the high Difference in Reactive Power for the fault samples 52 and 304.

Interestingly, the Minimum Power Factor and Difference in Reactive Power were not emerging as important features with the global interpretability approach, which is based on the weights magnitude of the linear models. Indeed, an approach that averages the contribution of the different factors across the whole dataset is likely to conceal the importance of

configurations in the features value that appears only in a few samples. On the other hand, by analyzing samples individually, IG can reveal new patterns in the data and allow to gain deeper insights about the true causes underlying specific faults.

VI. CONCLUSION

In this work, we tackled the problem of detecting unscheduled faults in the power grid, which have major consequences for customers, such industries, relying on a stable power supply. In collaboration with the DSO, we built a data set consisting of meteorological and power data variables, which monitor potentially relevant factors to cause power faults. Once the dataset was constructed, we trained different classifiers to detect imminent faults from the value of meteorological and power variables.

The classification performance was compared in terms of F1 score and the MLP classifier achieved the top performance, followed by the Ridge Classifier. The good classification results motivated the interpretation of the decision process learned by the model, as a tool to identify the variables that mostly explain the onset of power faults. We explored two different interpretability techniques. First, we considered the magnitude of the coefficient of the linear models to quantify the importance that, on average, the different features have to determine if a sample in the dataset is a fault. The results indicated that the amount of Flicker and Wind speed of gust are the most important variables in explaining the power disturbances. Such a global interpretability approach averages the contribution of the different factors across the whole dataset and, therefore, might fail to show interesting configurations in the features value that appear only in a few samples.

As a second interpretability technique, we used the Integrated Gradients to interpret the decision process taken by the MLP classifier on individual samples. This second approach allowed us to understand what features were considered important to classify a specific sample as a fault. Interestingly, some samples were classified as faults not only for having high values in Flicker and Wind speed of gust. In fact, the IG technique showed that the MLP classified as faults samples where the Minimum Power Factor was below average or where the Difference in Reactive Power was higher than average.

The proposed interpretability techniques revealed important patterns in the data, which allow us to gain deeper insights into the underlying causes of power faults. This type of knowledge will help the DSO to give more reliable warnings to its customers (both producers and consumers) that there is an enhanced risk for grid failure when certain meteorological and power flow conditions are met. With this information, the customers can take preemptive actions to reduce the negative consequences occurring when a fault strikes.

By understanding the major factors causing the faults, the DSO will also be able to better plan how to strengthen the grid to withstand incoming faults. Typical actions to improve the grid stability are: i) make changes in grid topology, such as optimizing coupling to make the grid stronger, isolating parts of the grid more likely to fail, running island mode whenever possible; ii) optimizing or even increasing the local production by introducing new power sources, including renewable ones; iii) reduce or adjust power flows by controlling flexible loads.

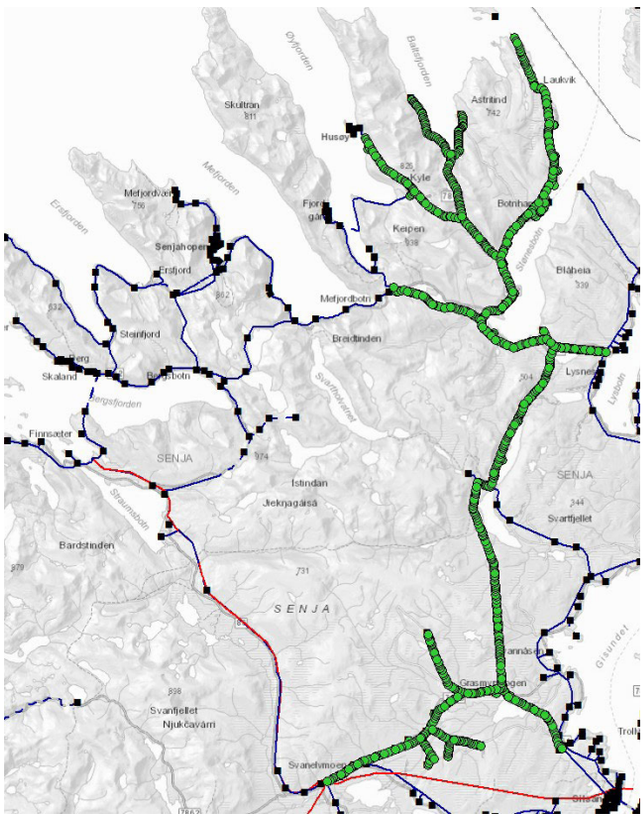


FIGURE 8. The SVAN22LY1 power grid. The power is distributed towards the north from the south. Each green dot represents a unique position of a utility pole.

These kinds of strategies to mitigate incoming grid faults are currently being developed by the DSO operating the grid in our study. In particular, the local power company is installing a large battery system that should be activated right before an incoming power fault, to provide additional power and avoid instability in the power supply. Understanding which variable should be monitored to detect an incoming power fault is, therefore, fundamental to plan the installation and management of the batteries.

APPENDIX A
THE INVESTIGATED POWER GRID

The power grid analyzed in this study is a radial distribution system serving an Arctic community located approximately at (69.257°N, 17.589°E). Arva Power Company, the DSO of the power grid, has named this specific grid as SVAN22LY1. Fig. 8 shows an overview of the whole SVAN22LY1 grid, indicated by green dots. The SVAN22LY1 grid spans over 60 kilometers from the south to the northernmost point and

has several branches to various communities towards the north. There are 978 unique utility poles (marked by green dots in Fig. 8) that support the power lines. The black boxes in Fig. 8 represent the electric transformer stations connected to the power grids. The red lines represent a power grid with an operating voltage of 66 kV, while the blue lines represent a power grid with an operating voltage of 22 kV. The SVAN22LY1 radial grid covered by green dots has an operating voltage of 22 kV. The largest customers connected to the SVAN22LY1 grid are located at the end of the northernmost point of the radial.

The total energy demand in the SVAN22LY1 grid is dominated by the load consumption of the local industry. The industry performs fish processing activities that are highly seasonal and uses many electrical machines in the production line that require stable power quality. Even minor power disturbances in the power supply trigger significantly long interruptions since the automated production line needs to be reset. In particular, for every short-term power interruption that occurs, is necessary to wait from 40 minutes to 1 hour before resuming the production. The consequences of the power disturbances are exacerbated by the topology of the power grid, which has a radial distribution with no alternative power supply in periods with disturbances.

APPENDIX B
DATASET CONSTRUCTION

A. FAULT REPORTS

The reported faults used in this study are logged by a power-quality (PQ) metering system, which was installed in February 2021 in the proximity of the local industries to continuously measure the power quality.

The PQ meter installed by the DSO is Dranetz HDPQ-DN, and is a monitoring instrument that is used for continuous monitoring of power systems. Such PQ meters are valuable tools to get better insight and knowledge about the actual power quality. More technical details about the PQ meter are provided in [58]. The PQ metering system reports all incidents with a voltage variation of $\pm 10\%$ from the nominal values on each phase of a three-phase system with phases A, B, C. According to the standard definition, all variations of $\pm 10\%$ from normal conditions are defined as a voltage variation, and a drop larger than 10% is referred to as a voltage dip [59]. Voltage dips could provoke tripping of sensitive components such as industrial machines.

B. WEATHER MEASUREMENTS

The weather variables that are considered relevant in causing power faults are: wind speed of gust, wind direction, temperature, pressure, humidity, and precipitation. The weather data are collected from areas that are more exposed to harsh weather conditions, such as hills and cliffs near the sea coast. To collect the weather data in the Arctic region of interest, we used the AROME-Arctic weather model .[¶] This model is developed by the meteorological institute of Norway (MET) and provides a reanalysis of historical weather data since November 2015 with a spatial resolution of 2.5 kilometers and a temporal resolution of 1 hour.

To collect the weather variables, the geographical coordinates from the weather-exposed areas in the power grid are used as inputs to the AROME-Arctic model. The output from the AROME-Arctic model is a dataset of 6 weather variables from the weather-exposed areas that are analyzed.

C. ELECTRICITY LOAD MEASUREMENTS

It is reasonable to assume that some types of fault are not caused by weather phenomena but originate from external factors that influence the power flows on the grid. To capture these effects, 6 different power-related variables from the largest industry connected to SVAN22LY1 are collected. The variables selected as relevant to explain power faults are: difference in frequency, voltage imbalance, the difference in active and reactive power, minimum power factor, and, finally, the amount of flicker in the system. All variables are metered on three different phases (phases A, B, and C).

A *change in power frequency* could be caused if there is an imbalance between energy production and consumption in the system. If there is a change in the power frequency (50 Hz is the normal frequency), the imbalance could cause power disturbances for the end-use customers.

Voltage imbalance is a voltage variation in the power system in which the voltage magnitudes or the phase angle between the different phases are not equal. It is believed that rapid changes (big changes within seconds/minutes) in power consumption at large industries could affect the power quality. Therefore, the *difference in active and reactive power* for each phase within each minute is computed. If the difference is large, there is a high activity at the industries, which are reported by the locals to result in a larger probability for faults.

The *minimum power factor* represents the relationship between the amount of active and reactive power in the system. If the minimum power factor is low, there is an increased amount of reactive power in the system. In the end, the amount of flicker in the system is collected.

Flicker is considered as a phenomenon in the power system and is closely connected to voltage fluctuations over a certain time frame [60]. A voltage fluctuation is a regular change in voltage that happens when the machinery that requires a high load is starting. In addition, rapid changes in load demand could cause voltage fluctuations. If there are several start-up situations, or the load varies significantly during a given time frame, it will be measured a high amount of flicker in the system. The amount of flicker is particularly relevant in the industry considered in this study, as they have several large machines that require high loads and have a cyclical varying load pattern. In this study, the time frame of the flicker is 10-minutes, which is the standard for measuring the short-term flicker [7].

[¶] https://www.met.no/en/projects/The-weather-model-AROME-Arctic

The PQ metering system has a 1-minute resolution, while the weather data have a 1-hour resolution. To align the temporal resolution of the different types of variables, the power consumption data are sub-sampled by taking one sample every 60. As an alternative sub-sampling technique, we tested taking the average of the values within each batch of 60 consecutive samples of power measurements. However, the results did not change significantly and, therefore, the former sub-sampling method was adopted.

APPENDIX C
A BRIEF HISTORY OF EXPLAINABILITY IN DEEP LEARNING

Due to the presence of many non-linear transformations, it is difficult to interpret the decision process of a neural network. During the last decade, considerable research effort has been devoted towards developing insights into what a neural network learns and how it makes its decisions. While most of the explanatory techniques were originally developed in the field of computer vision, some of them can be applied also to neural networks that process sequential or vectorial data.

Gradient-based approaches aim at identifying which inputs have the most influence on the model scoring function for a given class. The pioneering work of Simonyan at al. [53] proposed to compute a saliency map by taking the gradient of the class activation score (usually, the input to the last softmax) with respect to each input feature. The visualization of the saliency maps was successively improved by using tricks such as clipping the gradients, averaging the gradients after adding Gaussian noise to the original images, and taking the absolute value of the gradients [54].

In [55], the authors propose a method to project the activations of an intermediate hidden layer back to the input space. The procedure consists in approximately inverting the operations of a CNN (affine transformations, ReLU activations, MaxPooling) from the hidden layer to the input layer. The result gives an insight into which details the hidden layer has captured from the input image.

The Guided Back Propagation approach performs the standard gradient back-propagation but, when a ReLU is encountered, the gradient is back-propagated only if both the gradient and the ReLU activation in the forward pass are positive [56].

As a drawback, gradient-based methods attribute zero contribution to inputs that saturate the ReLU or MaxPool. To capture such shortcomings, a formal notion of explainability (or relevance) was introduced in [51]. In particular, the authors introduced an axiom on the *conservation of total relevance*, which states that the sum of the relevance of all pixels must equal the class score of the model. The authors propose to distribute the total relevance of the class score to the input features with a method called Layer-wise Relevance Propagation (LRP). The class score is computed as the difference between the score obtained by the actual input and the score obtained by an uninformative input, called *baseline*. Each time the relevance is passed down from a neuron to the contributing neurons in the layer below, the total relevance of contributing neurons is preserved. All incoming relevances to a neuron from the layer above are collected and summed up before passing down further. By doing this recursively from layer to layer, the input layer is eventually reached, which gives the relevance of each input feature. The relevance of a neuron to its contributing inputs can be distributed based on the magnitude of the weights of the neural network layers.

While LRP followed the conservation axiom, it did not formalize how to distribute the relevance among the input features. To address this problem DeepLiFT [52] enforces an additional axiom on how to propagate the relevance down, by following the chain rule like gradients.

REFERENCES

[1] S. Chiaradonna, F. Di Giandomenico, and G. Masetti, "Analyzing the impact of failures in the electric power distribution grid," in *Proc. 7th Latin-American Symp. Dependable Comput. (LADC)*, Oct. 2016, pp. 99–108.

[2] C. Klinger, O. Landeg, and V. Murray, "Power outages, extreme events and health: A systematic review of the literature from 2011–2012," *PLoS Currents*, to be published.

[3] T. H. Meles, "Impact of power outages on households in developing countries: Evidence from Ethiopia," *Energy Econ.*, vol. 91, Sep. 2020, Art. no. 104882.

[4] M. Shuai, W. Chengzhi, Y. Shiwen, G. Hao, Y. Jufang, and H. Hui, "Review on economic loss assessment of power outages," *Proc. Comput. Sci.*, vol. 130, pp. 1158–1163, May 2018.

[5] S. Tully, "The human right to access electricity," *Electr. J.*, vol. 19, no. 3, pp. 30–39, Apr. 2006.

[6] G. S. Gopinath and M. Meher, "Electricity a basic need for the human beings," *AIP Conf. Proc.*, vol. 1992, no. 1, 2018, Art. no. 040024.

[7] N. Lovdata. *Forskrift Om Leveringskvalitet I Kraftsystemet*. [Online]. Available: https://lovdata.no/dokument/SF/forskrift/2004-11-30-1557

[8] C. Rubí. *The Challenges of Upgrading the Power Grid for a Decarbonised Electric Future*. [Online]. Available: https://informaconnect.com/the-challenges-of-upgrading-the-power-grid-f%or-a-decarbonised-electric-future/

[9] N. Sapountzoglou, J. Lago, B. De Schutter, and B. Raison, "A generalizable and sensor-independent deep learning method for fault detection and location in low-voltage distribution grids," *Appl. Energy*, vol. 276, Oct. 2020, Art. no. 115299.

[10] K. Chen, J. Hu, Y. Zhang, Z. Yu, and J. He, "Fault location in power distribution systems via deep graph convolutional networks," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 1, pp. 119–131, Jan. 2020.

[11] H. Khorasgani, A. Hasanzadeh, A. Farahat, and C. Gupta, "Fault detection and isolation in industrial networks using graph convolutional neural networks," in *Proc. IEEE Int. Conf. Prognostics Health Manage. (ICPHM)*, Jun. 2019, pp. 1–7.

[12] R. M. G. Ferrari, T. Parisini, and M. M. Polycarpou, "Distributed fault detection and isolation of large-scale discrete-time nonlinear systems: An adaptive approximation approach," *IEEE Trans. Autom. Control*, vol. 57, no. 2, pp. 275–290, Feb. 2012.

[13] E. Balouji, I. Y. H. Gu, M. H. J. Bollen, A. Bagheri, and M. Nazari, "A LSTM-based deep learning method with application to voltage dip classification," in *Proc. 18th Int. Conf. Harmon. Quality Power (ICHQP)*, May 2018, pp. 1–5.

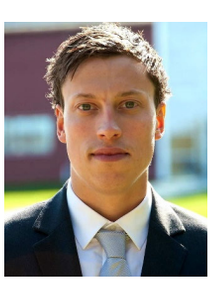
[14] F. M. Bianchi, E. De Santis, A. Rizzi, and A. Sadeghian, "Short-term electric load forecasting using echo state networks and PCA decomposition," *IEEE Access*, vol. 3, pp. 1931–1943, 2015.

[15] F. M. Bianchi, E. Maiorino, M. C. Kampffmeyer, A. Rizzi, and R. Jenssen, "Recurrent neural networks for short-term load forecasting: An overview and comparative analysis," Tech. Rep., 2017.

[16] A. T. D. Perera, V. M. Nik, D. Chen, J.-L. Scartezzini, and T. Hong, "Quantifying the impacts of climate change and extreme climate events on energy systems," *Nature Energy*, vol. 5, no. 2, pp. 150–159, Feb. 2020.

[17] H. Sabouhi, A. Doroudi, M. Fotuhi-Firuzabad, and M. Bashiri, "Electrical power system resilience assessment: A comprehensive approach," *IEEE Syst. J.*, vol. 14, no. 2, pp. 2643–2652, Jun. 2020.

[18] D. N. Trakas, M. Panteli, N. D. Hatziargyriou, and P. Mancarella, "Spatial risk analysis of power systems resilience during extreme events," *Risk Anal.*, vol. 39, no. 1, pp. 195–211, Jan. 2019.
[19] M. Panteli and P. Mancarella, "Influence of extreme weather and climate change on the resilience of power systems: Impacts and possible mitigation strategies," *Electr. Power Syst. Res.*, vol. 127, pp. 259–270, Oct. 2015.
[20] F. de Caro, E. M. Carlini, and D. Villacci, "Flexibility sources for enhancing the resilience of a power grid in presence of severe weather conditions," in *Proc. AEIT Int. Annu. Conf. (AEIT)*, Sep. 2019, pp. 1–6.
[21] D. Owerko, F. Gama, and A. Ribeiro, "Predicting power outages using graph neural networks," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Nov. 2018, pp. 743–747.
[22] O. F. Eikeland, F. M. Bianchi, I. S. Holmstrand, S. Bakkejord, S. Santos, and M. Chiesa, "Uncovering contributing factors to interruptions in the power grid: An Arctic case," Tech. Rep., 2021.
[23] O. P. Mahela, A. G. Shaik, and N. Gupta, "A critical review of detection and classification of power quality events," *Renew. Sustain. Energy Rev.*, vol. 41, pp. 495–505, Jan. 2015.
[24] M. Mishra, "Power quality disturbance detection and classification using signal processing and soft computing techniques: A comprehensive review," *Int. Trans. Electr. Energy Syst.*, vol. 29, no. 8, Aug. 2019, Art. no. e12008.
[25] K. Michalowska, V. Hoffmann, and C. Andresen, "Impact of seasonal weather on forecasting of power quality disturbances in distribution grids," in *Proc. Int. Conf. Smart Energy Syst. Technol. (SEST)*, Sep. 2020, pp. 1–6.
[26] F. Yang, D. W. Wanik, D. Cerrai, M. A. E. Bhuiyan, and E. N. Anagnostou, "Quantifying uncertainty in machine learning-based power outage prediction model training: A tool for sustainable storm restoration," *Sustainability*, vol. 12, no. 4, p. 1525, Feb. 2020.
[27] A. M. Salman and Y. Li, "A probabilistic framework for multi-hazard risk mitigation for electric power transmission systems subjected to seismic and hurricane hazards," *Struct. Infrastruct. Eng.*, vol. 14, no. 11, pp. 1499–1519, Nov. 2018.
[28] A. M. Salman and Y. Li, "A probabilistic framework for seismic risk assessment of electric power systems," *Proc. Eng.*, vol. 199, pp. 1187–1192, Jan. 2017.
[29] A. M. Salman and Y. Li, "Multihazard risk assessment of electric power systems," *J. Struct. Eng.*, vol. 143, no. 3, Mar. 2017, Art. no. 04016198.
[30] S. Mukherjee, R. Nateghi, and M. Hastak, "A multi-hazard approach to assess severe weather-induced major power outage risks in the U.S.," *Rel. Eng. Syst. Saf.*, vol. 175, pp. 283–305, Jul. 2018.
[31] R. Eskandarpour and A. Khodaei, "Machine learning based power grid outage prediction in response to extreme events," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 3315–3316, Jul. 2017.
[32] K. Murray and K. R. W. Bell, "Wind related faults on the GB transmission network," in *Proc. Int. Conf. Probabilistic Methods Appl. Power Syst. (PMAPS)*, Jul. 2014, pp. 1–6.
[33] V. Hoffmann, K. Michałowska, C. Andresen, and B. N. Torsæter, "Incipient fault prediction in power quality monitoring," Tech. Rep., 2019.
[34] K. W. Hoiem, V. Santi, B. N. Torsater, H. Langseth, C. A. Andresen, and G. H. Rosenlund, "Comparative study of event prediction in power grids using supervised machine learning methods," in *Proc. Int. Conf. Smart Energy Syst. Technol. (SEST)*, Sep. 2020, pp. 1–6.
[35] G. M. Abusdal, G. T. Heydt, and A. Ripegutu, "Utilization of advanced metering infrastructure in back-fed ground fault detection," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Jul. 2015, pp. 1–5.
[36] T. S. Tyvold, B. Nybakk Torsater, C. A. Andresen, and V. Hoffmann, "Impact of the temporal distribution of faults on prediction of voltage anomalies in the power grid," in *Proc. Int. Conf. Smart Energy Syst. Technol. (SEST)*, Sep. 2020, pp. 1–6.
[37] G. H. Rosenlund, K. W. Hoiem, B. N. Torsater, and C. A. Andresen, "Clustering and dimensionality-reduction techniques applied on power quality measurement data," in *Proc. Int. Conf. Smart Energy Syst. Technol. (SEST)*, Sep. 2020, pp. 1–6.
[38] R. Fainti, M. Alamaniotis, and L. H. Tsoukalas, "Three-phase line overloading predictive monitoring utilizing artificial neural networks," in *Proc. 19th Int. Conf. Intell. Syst. Appl. Power Syst. (ISAP)*, Sep. 2017, pp. 1–6.
[39] Y. Zhou, R. Arghandeh, and C. J. Spanos, "Partial knowledge data-driven event detection for power distribution networks," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 5152–5162, Aug. 2018.
[40] K. Manivinnan, C. L. Benner, B. D. Russell, and J. A. Wischkaemper, "Automatic identification, clustering and reporting of recurrent faults in electric distribution feeders," in *Proc. 19th Int. Conf. Intell. Syst. Appl. Power Syst. (ISAP)*, Sep. 2017, pp. 1–6.
[41] S. Khokhar, A. A. M. Zin, A. P. Memon, and A. S. Mokhtar, "A new optimal feature selection algorithm for classification of power quality disturbances using discrete wavelet transform and probabilistic neural network," *Measurement*, vol. 95, pp. 246–259, Jan. 2017.
[42] O. Zyabkina, M. Domagk, J. Meyer, and P. Schegner, "A feature-based method for automatic anomaly identification in power quality measurements," in *Proc. IEEE Int. Conf. Probabilistic Methods Appl. Power Syst. (PMAPS)*, Jun. 2018, pp. 1–6.
[43] R. A. Shuvro, P. Das, M. M. Hayat, and M. Talukder, "Predicting cascading failures in power grids using machine learning algorithms," in *Proc. North Amer. Power Symp. (NAPS)*, Oct. 2019, pp. 1–6.
[44] C. M. Bishop, "Pattern recognition," *Mach. Learn.*, vol. 128, no. 9, 2006.
[45] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proc. 5th Annu. Workshop Comput. Learn. Theory (COLT)*, 1992, pp. 144–152.
[46] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
[47] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
[48] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
[49] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, 2014, pp. 1–15.
[50] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic attribution for deep networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 3319–3328.
[51] S. Bach, A. Binder, G. Montavon, F. Klauschen, K.-R. Müller, and W. Samek, "On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation," *PLoS ONE*, vol. 10, no. 7, Jul. 2015, Art. no. e0130140.
[52] A. Shrikumar, P. Greenside, and A. Kundaje, "Learning important features through propagating activation differences," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 3145–3153.
[53] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps," in *Proc. Int. Conf. Learn. Represent.*, 2014, pp. 1–8.
[54] D. Smilkov, N. Thorat, B. Kim, F. Viégas, and M. Wattenberg, "Smoothgrad: Removing noise by adding noise," in *Proc. Int. Conf. Mach. Learn.*, 2017.
[55] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Eur. Conf. Comput. Vis.*, Springer, 2014, pp. 818–833.
[56] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for simplicity: The all convolutional net," in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 1–14.
[57] A. Mansour, Z. Chengning, and H. Nasry, "Measurement of power components in balanced and unbalanced three-phase systems under nonsinusoidal operating conditions by using IEEE standard 1459–2010 and Fourier analysis," in *Proc. Int. Conf. Technological Adv. Electr., Electron. Comput. Eng. (TAEECE)*, May 2013, pp. 166–171.
[58] D. Technologies. *Permanent PQ and Energy Monitoring Products*. [Online]. Available: https://www.dranetz.com/product/hdpq-dn/
[59] E. Csanyi. *Detailed Overview of Power System Disturbances (Causes and Impacts)*. [Online]. Available: https://electrical-engineering-portal.com/detailed-overview-of-power-sy%stem-disturbances-causes-and-impacts
[60] A. C. S. Mark Halpin, *Power Electronics Handbook*, 2nd ed. Montgomery, AL, USA: Academic, Auburn Univ., Department of Electrical and Computer Engineering, 2007, ch. 38.



ODIN FOLDVIK EIKELAND received the M.S. degree in energy, climate, and environment from the University of Tromsø, in 2019, where he is currently pursuing the Ph.D. degree. He is also a Civil Engineer in renewable energy, and took a Ph.D. related to renewable energy and energy analytics. His main research interests include renewable energy with a special focus on solar energy, time-series forecasting, neural networks, and power grid fault detection with machine learning techniques.

INGA SETSÅ HOLMSTRAND received the M.Sc. degree from the Department of Physics and Technology, University of Tromsø, in 2018. She started at Arva Power Company, in 2020, where she currently works with data analysis of the power grid.

SIGURD BAKKEJORD received the M.Sc. degree in electrical power engineering from the Norwegian University of Science and Technology, in June 2011. He is currently working as a Power Grid System Manager of the Distribution System Operator Arva AS. His main research interests include power grid analysis and grid-information systems.

MATTEO CHIESA is currently a Material Scientist working on creating and implementing technologies necessary to adapt the current energy system into a competitive and secure one. He design and investigate engineered nanomaterials for solar energy conversion and storage systems. His main research interest includes tackling specific challenges related to the deployment of renewable energy under arctic conditions through synthetic and symbiotic approach.

FILIPPO MARIA BIANCHI received the Ph.D. degree from the Department of Information Engineering, Electronics, and Telecommunications, Sapienza University of Rome, in 2016. He is currently an Associate Professor with the Department of Mathematics and Statistics, UiT-The Arctic University of Norway, and also a Research Scientist at NORCE. His research interests include machine learning, complex networks, and dynamical systems.

8

Paper III

Energy Conversion and Management: X 15 (2022) 100239

Contents lists available at ScienceDirect

Energy Conversion and Management: X

journal homepage: www.sciencedirect.com/journal/energy-conversion-and-management-x

Probabilistic forecasts of wind power generation in regions with complex topography using deep learning methods: An Arctic case

Odin Foldvik Eikeland [a], Finn Dag Hovem [b], Tom Eirik Olsen [b], Matteo Chiesa [a], Filippo Maria Bianchi [c,d,*]

[a] Department of Physics and Technology, UiT-the Arctic University of Norway, 9037 Tromsø, Norway
[b] Ishavskraft Power Company, 9024 Tromsø, Norway
[c] Department of Mathematics and Statistics, UiT-the Arctic University of Norway, 9037 Tromsø, Norway
[d] NORCE Norwegian Research Centre AS, Norway

ARTICLE INFO

Keywords:
Energy analytics
Probabilistic forecasting
Wind power electricity generation
Deep learning

ABSTRACT

The energy market relies on forecasting capabilities of both demand and power generation that need to be kept in dynamic balance. Nowadays, contracts and auctions of renewable energy in a liberalized electricity market heavily rely on forecasting future power generation. The highly intermittent nature of renewable energy sources increases the uncertainty about future power generation. Since point forecast does not account for such uncertainties, it is necessary to rely on probabilistic forecasts.

This work first introduces probabilistic forecasts with deep learning. Then, we show how deep learning models can be used to make probabilistic forecasts of day-ahead power generation from a wind power plant located in Northern Norway. The performance, in terms of the quality of the prediction intervals, is compared to different deep learning models and sets of covariates.

The findings show that the accuracy of the predictions improves by 37% when historical data on measured weather and numerical weather predictions (NWPs) were included as exogenous variables. In particular, historical data allows the model to auto-correct systematic biases in the NWPs. Finally, we observe that when using only NWPs or only measured weather as exogenous variables, worse performances are obtained. The work shows the importance of understanding which variables must be included to improve the prediction performance, which is of fundamental value for the energy market that relies on accurate forecasting capabilities.

1. Introduction

Making accurate predictions is of fundamental importance in the energy market where decisions are based on expectations about the future. Today, when it comes to renewable energy generation, such decisions are increasingly made in a liberalized electricity market environment, where future power generation must be offered through contracts and auction mechanisms, hence based on forecasts [1–3]. Since renewable energy sources (RES) are to eventually participate in market mechanisms under the same rules as conventional fossil-fuel-based generators, mismatches between contracted generation and actual deliveries may induce financial penalties [4]. Indeed, the energy production from RES can be predicted with limited accuracy. This, in addition to uncertainties in market prices, yield uncertain market returns. However, in market environments under such high levels of uncertainty, the relevant stakeholders may make better decisions if they are given the best possible estimates about the future.

Forecasts in their most common form are to predict the next value that is most likely to occur. Forecasts in such form are called a *point* forecast. The main objective of making point forecasts is to train a model to predict a certain point in the future, and hopefully, the actual value will eventually be the same (i.e., a perfect prediction). However, it is not reasonable to expect that a model can perform 100% accurate predictions. The models will therefore predict the most likely value based on what it has learned through the information that is given. However, when predicting the most likely value (or point) in the future, one does not consider the uncertainties in this forecast (how sure are we that this will be the next value). For the stakeholders that work in the energy market with a high degree of uncertainty, it is of interest to measure the uncertainty in a given forecast. For instance, the power production from

* Corresponding author.
E-mail address: filippo.m.bianchi@uit.no (F.M. Bianchi).

https://doi.org/10.1016/j.ecmx.2022.100239

Available online 27 May 2022

solar photovoltaic (PV) or wind power is highly intermittent and dependent on multiple features with a complex and non-linear nature (weather, power market, human activity) [5–7,1]. Making accurate point predictions from these technologies is almost impossible, especially when considering short-term forecasts that range from 12-h to 36-h ahead, which is the day-ahead market the energy trading companies must relate to [6,8]. Therefore, it is of fundamental importance to be able to predict the *distribution* of the expected outcome to find a certain interval of possible outcomes. Probabilistic forecasts predict such distributions where the interval of possible outcomes creates a prediction interval (PI). Probabilistic forecasts will allow market participants to consider the uncertainties in a given prediction.

The increased share of RES technologies with fluctuating generation in the electricity market, and the rapid development within machine learning (ML), have motivated the development of research concerning probabilistic forecasts in energy applications with deep learning. Efforts have been done to construct deep learning models that create probabilistic forecasts of expected generation from technologies such as PV and wind power [2,9,7,10,11].

In works by Sadeghi, Saleh, et al. [12,13] a novel deep learning-based approach to model the uncertainty in a virtual power plant (VPP) concept was proposed. The VPP concept is used to address the challenges related to the increased penetration of RES and electric vehicles in the electricity system, which has highly intermittent nature in generation and consumption [14].

Similar approaches have been used to make probabilistic forecasts of expected weather (such as wind speed), traffic, energy load, and spot prices in the electricity market [15–17,6].

The majority of former works have used different deep learning architectures such as Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs) where different loss functions are optimized to compute predicted distributions. In addition, some works are proposing methodologies where original Neural Networks (NNs) are modified for a specific task to make probabilistic forecasts [18,8,10]. In [19] the authors compared deep learning models to make probabilistic forecasts on different datasets and showed that wind power generation was the most challenging one to predict with high accuracy due to the highly non-linear feature and randomness in power generation. Despite a large number of works on probabilistic forecasts with deep learning techniques in energy applications, there is a lack of research comparing the effect of using exogenous variables to obtain the best prediction performance of wind power generation.

The electricity generation from wind power systems is directly affected by the amount of wind that hits the turbines, where the amount of power that is generated follows a wind-to-power conversion process, or the *power curve* [20]. Therefore, the majority of work uses numerical weather predictions (NWPs) as inputs to forecast the expected wind power generation in the short-term range (6–72 h ahead) [21–24]. When predicting wind power generation, a large part of the forecast error will therefore directly come from the NWPs. Predicting short-term weather (especially wind speed) is a complex task due to its non-linear and fluctuating nature [25–27]. Consequently, this gives large sources of error for making accurate short-term predictions of wind power generation. The wind power case study considered in this work is in a region with complex terrain where it is large weather variations within small distances, making it increasingly difficult to have accurate NWPs available. In this work, we address the issue of inaccurate NWPs by constructing different configurations of datasets that could account for such failure sources. We perform prediction experiments using different sets of exogenous variables are compared to make probabilistic forecasts of wind power generation with deep learning models. The exogenous variables considered are the NWP (wind speed + wind direction) and measured weather (wind speed + wind direction) from instruments installed locally in the wind power plant. The wind power plant studied in this work lies in a region with a complex topography in Arctic weather conditions that potentially have a larger source for failure in the NWPs compared to other regions. Therefore it is a motivation in this work to use deep learning models that are trained to account for possible failure sources in weather predictions by using the measured weather to correct for such failures. In the end, we compare the accuracy of the probabilistic forecasts for different sets of covariates and discover the variables that should be included to compute high-quality PIs.

This paper is structured as follows. In Section 2, an introduction to the field of probabilistic forecasts is provided. Section 3 presents a review of relevant research in the field of probabilistic forecast within energy analytics. In Section 4, the wind power plant case study is presented. Section 5 presents the methodology for making probabilistic forecasts. In Section 6 the results are presented in terms of the obtained quality of the PIs. Conclusions are given in Section 7.

2. The concept of prediction intervals

Although the field of probabilistic forecasts is based on well-known concepts in statistics, and there exists a vast amount of existing research on this field, some concepts have been shown to be confused and misused [28]. To avoid this, we start by defining and explaining some relevant concepts that will be used through the study.

When making probabilistic forecasts, the goal is to make a PI that considers the uncertainties in the predictions. The PI is an estimate of an interval in which the future observation will fall with a certain probability. For instance, for a 95% PI, there should be a 95% probability that the next value will fall within the lower and upper bounds.

The generic form for computing the upper and lower bound of PIs is

$$P(Y_n \in C(X_n)) \geqslant 1-\alpha, \tag{1}$$

where Y_n is the response variable, $C(X_n)$ is the confidence interval centered on the covariate X_n, and α is the significance level. For a 95% PI, α is 0.05.

PIs can either be of low quality, or of high quality. A high-quality PI will be sharp and well-calibrated (the PI contains the desired proportion of the samples $1-\alpha$). A low-quality PI is too wide or is miscalibrated (the empirical quantiles do not match the theoretical ones).

Examples of high-quality and low-quality PIs is given in Fig. 1.

Here, we see an example of a point-and probabilistic forecast. The red dots represent the observed values, the blue dots are the point predictions. The yellow line represents the PI, which in this case is the 95% prediction interval. Here, it is clear that the high-quality PI (uppermost Figure) fulfills two important criteria when computing PIs. Sharpness and calibration. The boundaries of the PI should be sharp to convey useful information about the uncertainity in the predictions. In addition, the PI should be calibrated, which means that for a 95% PI, 95% of the observed values should fall within the PI. Here, the PI contains exactly 95% of the observed values, and hence we can trust that there is a 95% probability that the observed values will fall within this PI.

The middle and lowermost examples on contrary, show examples of low-quality PIs. The example in the middle shows a PI with high sharpness, but it is not calibrated as it only contains 80% of the observed values, which is not adequate in a 95% PI. The lowermost exampled shows a calibrated PI as it contains 95% of the observed values. However, it is not sharp, and PIs that are too wide do not convey any useful information about the variation in the observed values.

3. Related works on probabilistic forecasting with deep learning

There exists a vast amount of former research on probabilistic forecast in different applications, and it is outside the scope to review all of them in this work. Therefore, former works that have particularly focused on the application of probabilistic forecast with deep learning within energy applications are reviewed here. In the following, a few examples of some former popular reviews that have high relevance for energy applications is provided.

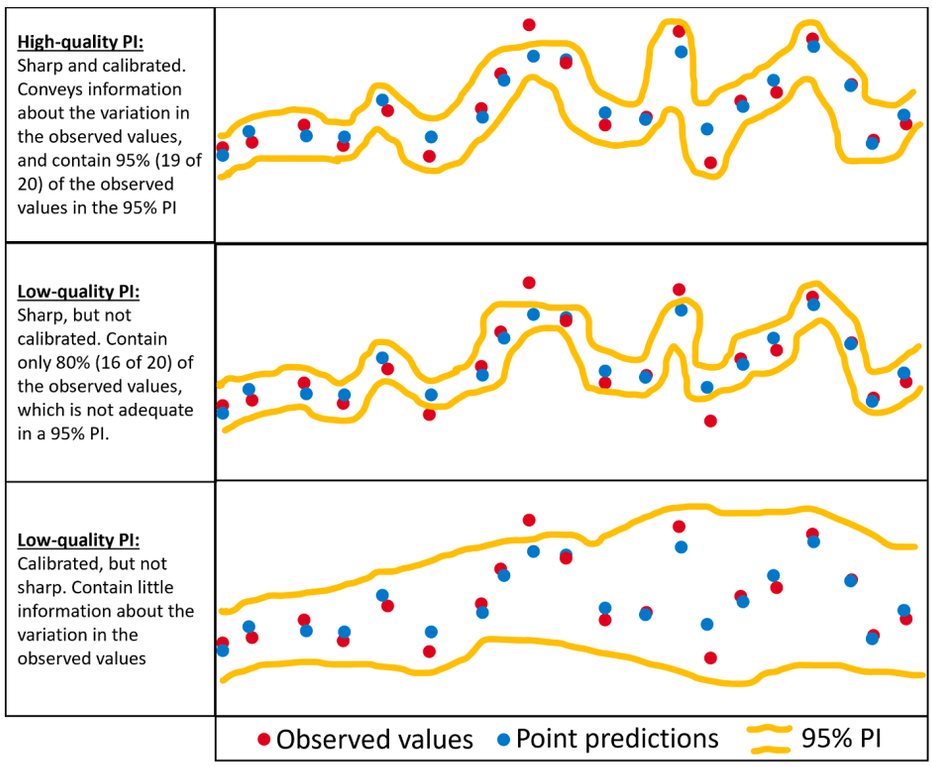


Fig. 1. Examples of PI with high (uppermost figure) and low quality. The red dots represent the observed values, the blue dots is the point predictions. The yellow line represent the PI, which in this case is the 95% prediction interval.

In [28], a comprehensive review on probabilistic forecast of the electric load was conducted. The authors offered a tutorial review of probabilistic forecasting and presented a brief introduction to common prediction horizon categories which are used in several works considering forecasting within energy applications [20,21]. The typical forecast-horizons range from very short (second to minutes), short (hours to days), medium (days to weeks), and long term (weeks to years). Very short-term predictions are often used for power wind turbine control, and for such applications, only historical measurement data has shown to be sufficient to make accurate predictions. Short-term predictions are popular to use for intraday and day-ahead electricity market purposes. Here, information from NWPs are required to improve the forecast accuracy [20]. Medium and long-term predictions are often used for maintenance scheduling, power system planning, and resource (solar or wind) assessment. In addition, the authors presented some notable techniques, methodologies, and evaluation methods that could be valuable for researchers and practitioners in the area of probabilistic load forecasting. To improve the field of probabilistic forecasts, they underlined the need to invest in additional research, and they state the importance of considering emerging technologies and energy policies in the probabilistic forecasting process.

Another relevant review was performed by [20]. The authors discuss the challenge of making accurate predictions of the highly intermittent electricity generation from wind power. They discuss the value of probabilistic forecasts in such tasks as it could provide additional quantitative information on the uncertainty. For actors that work with decision-making in an uncertain environment (for instance power traders), information about uncertainties in decisions is of interest. In this work, the authors presented a review of state-of-the-art methods and new developments within wind power probabilistic forecasting. They discuss different forecast methods and classified them into different categories in terms of uncertainty representation. Finally, they summarized requirements and the overall framework of uncertainty forecasting evaluation.

Notably, both these two famous reviews are quite mature (2016 and 2014), and the field of probabilistic forecasting with deep learning in energy analytics has vastly developed in recent years. A recent review and comparison study was performed by Mashlakov et al. [19]. Here, the recent developments in the field of probabilistic forecasting, multivariate models, and multi-horizon time series forecasting were reviewed. The authors empirically evaluate the performance of novel deep learning models for predicting wind power, solar power, electricity load, and wholesale electricity price for intraday and day-ahead time horizons. They reviewed the performances of both point and probabilistic forecast approaches. This comprehensive comparison work could serve as a reference point for the quantitative evaluation of deep learning models for probabilistic multivariate energy forecasting in power systems.

In addition to the above-mentioned reviews, Table 1 present an overview of some recent research works which have contributed to developing the field of probabilistic forecasting with deep learning in energy applications.

The works in Tab.1 are grouped according to which model they use to make either point-or probabilistic forecasts. The most common methodologies involve using either Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs).

The authors in [9] addressed the challenge of making accurate predictions of electric generation from intermittent photovoltaics. They proposed a two-step methodology where the first step was to use wavelet transform to decompose the signal into several frequencies, where each frequency has better behavior and outlines. Then, a deep convolutional neural network was employed to extract the nonlinear features and invariant structures exhibited in each frequency. The point of the hybrid approach is that the final decomposed signal is easier to predict than the original time series and should result in better prediction accuracy. To perform predictions, the deterministic method and a spline quantile regression were developed to evaluate the probabilistic forecasts. The proposed deterministic and probabilistic methods were applied to real-world datasets from PV farms in Belgium, and the results demonstrated improvement compared to more conventional forecasting methods. In

Table 1
Overview of different works.

Model	Year	Network	Forecasts	Metrics	Dataset
CNNs					
[9]	2019	CNN	Point, Prob.	MAPE, RMSE QL	PV power
[29]	2020	TCN	Prob, Point	QL, NRMSE	Retailers
[2]	2020	CNN	Prob.	RMSE, MAPE PCIP, PINAW	PV power
RNNs					
[30]	2019	RNN	Prob	QL, NRMSE	demand, traffic finance
[16]	2020	RNN	Prob.	QL, NRMSE	demand, traffic, electricity
[31]	2021	RNN	Prob	CRPS, MSE	traffic, energy
[32]	2020	NN library	Prob	QL	energy, PV
Modified NNs					
[8]	2018	Bi-LSTM	Prob	RMSE, QL	Power market
[15]	2020	BiGRU	Point, Prob.	RMSE, QL	Wind speed
[7]	2020	IDMDN	Prob	NRMSE	Wind power
[18]	2020	GRU + CNN	Point, Prob	RMSE, CRPS	Wind speed
[17]	2018	iQRNN	Prob.	QL, WS	Energy load
[33]	2020	sDAEs + ANN	Point	RMSE, MAPE MAE, R^2	Wind speed
[34]	2021	PCFM	Point, Prob	RMSE, MAE PINAW, PCIP	Wind speed
[35]	2019	WT + CNN	Point, Prob	RMSE, MAE, MAPE PINAW, PCIP, CWC	Wind speed
[36]	2018	NARX	Point, Prob	RMSE, MAE PINAW, PCIP	Wind speed
[37]	2019	LSTM + GPR	Point, Prob	RMSE, MAE, CRPS	Wind speed
[38]	2021	MC + Bi-LSTM	Point	MAE, MAPE, RMSE	Load demand Electricity price Wind speed
[38]	2020	SDAEs + SR-NNs	Point	MAE, MAPE, RMSE	Wind speed
Bayesian methods					
[6]	2019	Bayesian	Prob.	RMSE, MAPE CRPS	Energy prices
[1]	2020	STNN	Point, Prob.	RMSE, CRPS	Wind speed
[39]	2020	Bayesian NNs	Point, Prob	RMSE, CPRS	Energy load
[40]	2019	Calibrated	Prob.	MAPE, QL	UCI datasets
Ensemble methods					
[10]	2017	WT + CNN	Prob	RMSE, PICP	Wind power
[5]	2016	7 ML models	Prob	RMSE, MAE QL	Solar power
[11]	2021	SEFMGPR	Prob	RMSE, R^2	Wind power
[41]	2020	EPS	Prob	CRPS	Solar irradiance
[42]	2020	MDE	Prob	QL	Wind power

[29], a probabilistic forecasting framework were proposed. The framework is based on CNNs and can be applied to estimate the probability density under both parametric and non-parametric settings. The loss function in the non-parametric framework is obtained by quantile regression, while for the parametric approach, the negative log-likelihood function is constructed as the loss function. The authors stack residual blocs based on dilated causal convolutional nets to capture the temporal dependencies of the time series, which creates a temporal convolutional neural network. The framework is able to learn latent correlations among series and should handle complex real-world patterns such as seasonality and holidays. The proposed method shows a high degree of flexibility and could include exogenous covariates such as weather forecasts. The authors demonstrate the framework on several real-world datasets and show that the results outperformed the state-of-the-art in both point and probabilistic forecasting. In [2], a two-stage training strategy is proposed to optimize the training procedure of quantile CNN (QCNN). The approach is developed with the aim to tackle the challenging task of making accurate probabilistic forecasts of PV power generation which has a high degree of variability. The two-step approach is as follows: First, the QCNN constructs a feature extraction network based on CNNs to mine the deep features of the PV power influence factors. Thereafter, quantile regression is employed to generate the PV power probability distribution based on the extracted features. To tackle a common training problem of the QCNN, a CNN is trained using a deterministic forecasting method, and a QR is trained using a linear-programming method. This two-stage process ensures that the training problem of QCNN could be avoided. They demonstrate the method on a real-world PV power plant in Australia and show promising results compared to other forecasting methods.

Instead of using CNNs, some authors used RNNs to make probabilistic forecasts. In [30] they propose a flexible method for probabilistic modeling with conditional quantile functions using monotonic regression splines. They parameterize the shape of the spline using a neural network. The parameters in the neural network are learned by minimizing the CRPS. By following this approach, they propose a method for probabilistic time series forecasting by combining the modeling capacity of RNNs with a spline-based representation of the output distribution. In addition, by minimizing the CRPS, they avoid the quantile crossing problem. Finally, the proposed approach does not require the parametric assumptions made on the output distribution and the observed data. The vast majority of techniques assume Gaussian distribution, which is often not adequate in real-world datasets. The authors in [16] proposed a method for making probabilistic forecasts with autoregressive recurrent neural networks (DeepAR). The proposed method learns a global model from historical data of all the time series in the dataset and produces probabilistic forecasts. The main contribution of the proposed DeepAR was twofold. First, they proposed an RNN architecture for probabilistic forecasting by incorporating a Negative Binomial likelihood for count data as well as special treatment for the case when the magnitudes of the time series vary widely. Second, they demonstrate the results empirically on several real-world data sets and showed that this model produces accurate probabilistic forecasts across a wide range of input characteristics. In [31], the challenging task of probabilistic forecasts of high dimensional multivariate time series is tackled. They introduce a novel temporal latent auto-encoder method that enables nonlinear factorization of multivariate time series. This method offers an efficient combination between flexible non-linear autoencoder mapping and inherent latent temporal dynamics modeled by a Long Short Term Memory (LSTM) RNN. This method is learned end-to-end with a temporal deep learning latent space forecast model. By imposing a probabilistic latent space model, complex distributions of the input time series are modeled via the decoder. They demonstrate through several experiments that their proposed model achieved state-of-the-art performance on many popular multivariate datasets.

Besides the CNN and RNN architectures, several works have proposed approaches using Bi-directional LSTM or Bi-directional GRU models with promising results [8,15,38]. In addition, several works have proposed methodologies where original Neural Networks have been modified for the specific probabilistic forecasting task. We refer to the works that are listed in the "Modified NNs" section in Tab.1 for more detailed descriptions. Most works listed in Tab.1 have applied the frequentist approach where the main aim is to optimize a cost function (such as the quantile loss), but some works have also proposed a Bayesian approach to make probabilistic forecasts with deep learning (see the "Bayesian methods" section for more detailed descriptions).

The final approach for making probabilistic forecasts that are listed in Tab.1 is the ensemble method. Ensemble techniques are a method where multiple learning algorithms are combined to obtain better

prediction performance than could be obtained from a single learning algorithm alone. For instance, the authors in [5] tackled the challenging task of predicting electric power generation from PVs using seven different ML models that are ensembled together to compute probabilistic forecasts. The findings show the ensemble methods outperformed the individual ML models.

All works in Tab.1 have tested the performance of their proposed methodology on energy-related datasets, and the prediction performances are presented with popular probabilistic metrics such as the QL, CRPS, PCIP, and PINAW (some of these metrics are defined and used in the result section in this work). Noteworthy, among all works, there is no specific model or method that seems to consistently achieve the best results, apart from the DeepAR model which has shown promising results in several works [19,16,32].

In a recent review and comparison work by [19], the authors highlighted the need of testing more automated deep learning models to progress the research within the field of probabilistic forecasts in energy applications. The authors suggested using open source libraries for modeling, such as the GluonTS toolkit [32]. The aim of the GluonTS library developed by [32] is to provide a flexible tool for probabilistic time series modeling with deep learning-based models. Motivated by the suggestion in [19], we implement deep learning models from the GluonTS library to make probabilistic forecasts of wind power generation.

4. Wind power plant case study

4.1. The wind power plant located in Northern Norway

In this work, the power generation from a wind farm in the region of

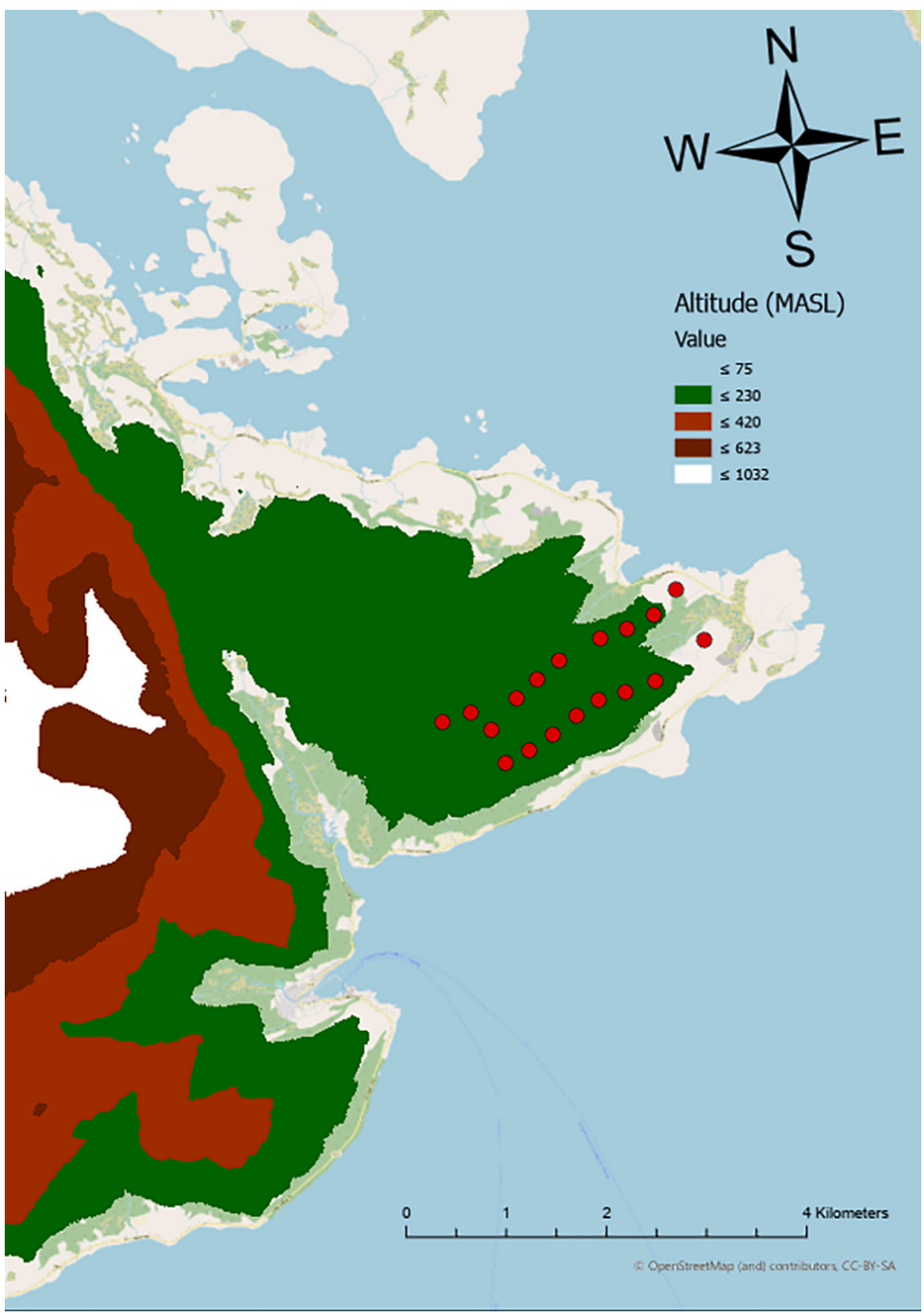


Fig. 2. Altitude map where the different colors indicate the altitude level. Green color is altitudes between 75 and 230 MASL. The highest mountains around in the region close to the wind park is up to 1032 MASL. The 18 turbines in the wind farm are marked in red dots.

Northern Norway is predicted. This wind farm consists of 18 turbines with a maximum power generation capacity of 3 MW, giving a maximum capacity of 54 MW for this power plant. In Fig. 2, an altitude map is created for the region of the wind power plant.

The different colors in the altitude map indicate the interval in Meter Above Sea Level (MASL). Here, the green color is all altitudes between 75 and 230 MASL. The highest elevation in the mountains to the right has an altitude between 632 and 1,032 MASL. The red dots are the geographical positions of the different turbines. Most of the turbines are approximately at the same altitude below 230 MASL. The two turbines at the rightmost positions are slightly below an altitude of 75 MASL.

Such a complex topography where the altitude varies from 0 to above 1000 MASL within small regions, is a typical phenomenon for the terrain in Northern Norway. This terrain has a huge impact on the local weather variation, and this particular wind power plant is highly affected by local variations in the wind. The owner of this wind power plant reported that the leftmost turbine produced 25% less energy than the rightmost turbine during the year 2020. The reason is that the turbines located to the left are shielded by the large mountains (indicated by the white color in the altitude map) which reduce the amount of wind that hits the turbine. On the other hand, the turbines located to the right are close to the ocean where there are few objectives that could reduce the wind. This power plant is a perfect example of where it can be large differences in production between the turbines. This can increase the difficulties in making predictions for the whole park, as the total power output from the park will be even more fluctuating compared to wind farms in flat regions where the weather conditions are more equal throughout the whole farm.

4.2. Dataset and and technical properties of wind turbines

The data available from the wind park is the historical measured power generation, wind speed, and wind direction measured on each turbine from the year 2020. The measured wind speed and wind direction variables are collected from the weather stations mounted on the nacelle on the turbine that is located in the middle of the wind park. The data are in a 1-h resolution and are received from the Troms Kraft Power company, the owner of the wind power plant. This gives a dataset of totally 8,784 samples. In addition to measured data, the NWP from the AROME-Arctic model is collected. This model is developed by the meteorological institute of Norway (MET) and provides weather forecasts with a spatial resolution of 2.5 km and a temporal resolution of 1 h. Similar to the weather measurements, the predicted wind speed and wind direction variables are collected from the AROME-Arctic model. Using only wind speed and wind direction as additional input variables to the prediction experiments has been shown to provide the most accurate forecasts in former wind power forecast literature [21]. This is due to the fact that wind power generation is directly affected by the amount of wind that hits the turbine blades. Hereafter, the weather data is referred to the wind speed and wind direction. Fig. 3 illustrate a map with the AROME-Arctic simulated wind speed for a randomly selected hour on the 1st of March in 2020. The colors in the map represent the simulated wind speed in each cell (spatial resolution of 2.5 km). It is clear that the turbines in the wind power park are distributed over two different cells, and there are differences in the wind speed for the two cells.

Fig. 3 show large differences in wind speed within a few kilometers, ranging from zero wind (blue color in southernmost part) to almost 30 m/s wind (red color in northernmost part). The AROME-Arctic weather forecast for this particular example shows that the rightmost part of the wind power plant has a wind of 5.5 m/s, and the leftmost part has a wind of 4.9 m/s. To consider possible local differences in wind speed and wind direction, the weather data from both cells in the AROME-Arctic weather simulation map is included as exogenous variables.

Before proceeding to the prediction methodology, some technical features of wind power generation must be highlighted. Wind power

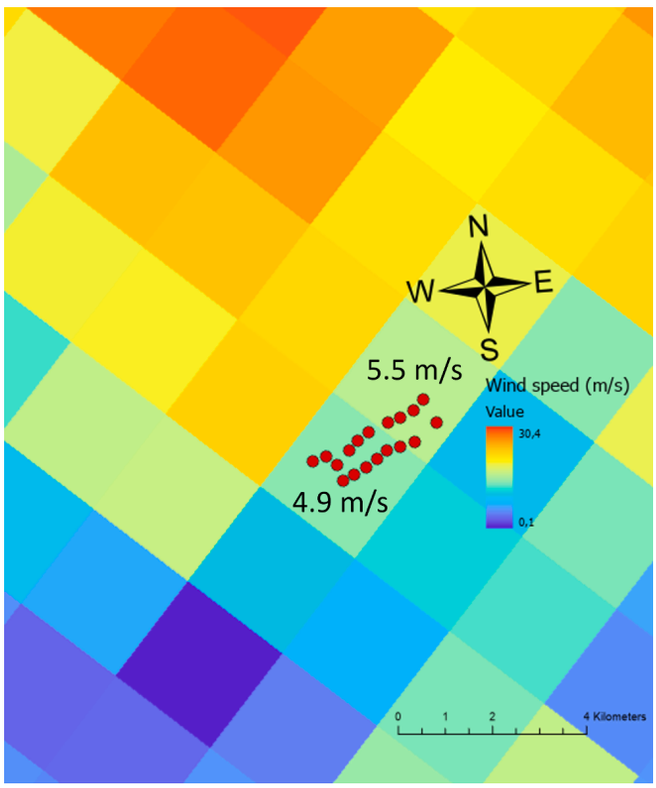


Fig. 3. The Arome Arctic weather simulation map where the red dots represent the position of each wind power turbine. Each square represent the spatial resolution of 2.5 × 2.5 km. 12 turbines and 6 turbines are located in the leftmost and rightmost cell, respectively.

generation is a (RES) technology that has a highly intermittent power generation due to the strong dependency on weather conditions. In addition, the power generation from the wind turbines is dependent on the *power curve*. The power curve from one of the turbines in the wind farm analyzed here is given in Fig. 4.

As seen in Fig. 4, the wind turbine has zero production between 0 m/s to 3/4 m/s. At the cut-in wind speed around 4 m/s, the power generation increases towards a wind speed of 12–13 m/s. At this wind speed, the power production reaches the maximum limit (or the rated power) of 3 MW around and produces at maximum it suddenly drops towards zero in the cut-off interval around 25 m/s. The reason for this drop is due to safety. If the blades at the turbine rotate too quickly, it can damage the equipment. Therefore, the turbines are forced to stop.

5. Methodology

5.1. Dataset configuration, preliminary analyses and deep learning models

In this work, the aim is to predict the day-ahead wind power generation. The day-ahead market closes at 12:00 where the market participants must submit their final bids to the electricity market about the expected amount of power generation the next day. Therefore, in this work, the forecast horizon is 36 h to consider all hours in the next day (12 h + 24 h).

First, in order to identify the time varying patterns in the time series, some statistical analyses are performed by computing the autocorrelation function (ACF) and the partial autocorrelation function (PACF). Fig. 5 shows the ACF and PACF graphs for the time series of the wind power generation and wind speed in the year of 2020. The ACF and PACF plots show no repetitive patterns, which indicate lack of seasonality in the time series. Interestingly, the time series show short-term

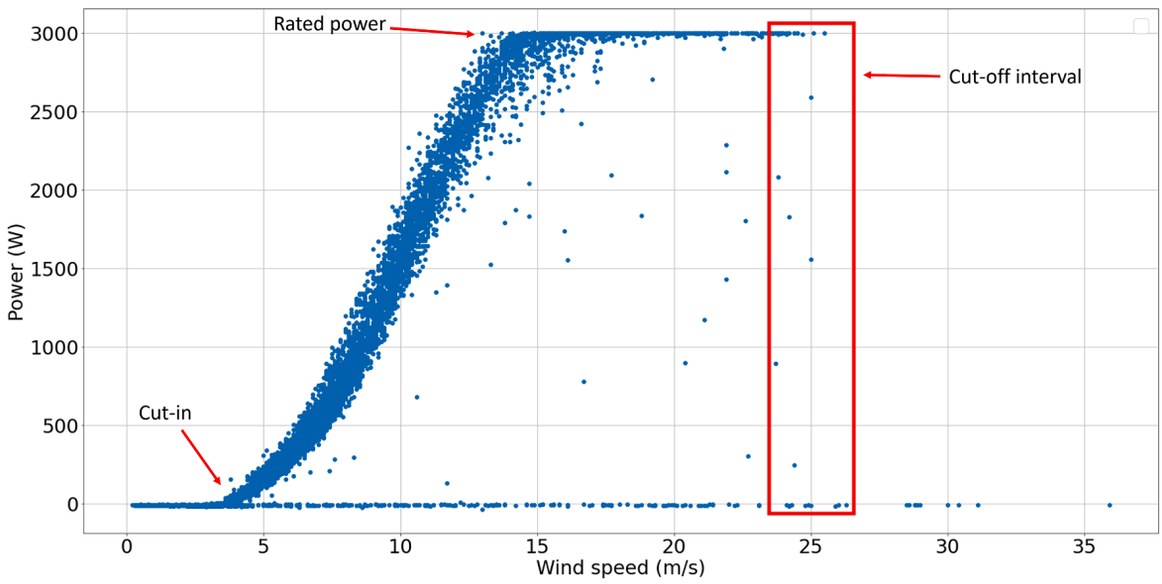


Fig. 4. Power curve from a wind turbine in the wind farm studied in this work with the cut-in, rated and cut-off wind speed marked. The zero production above 4 and below 25 m/s wind represent periods where the turbine is shut down due to turbine failure or maintenance.

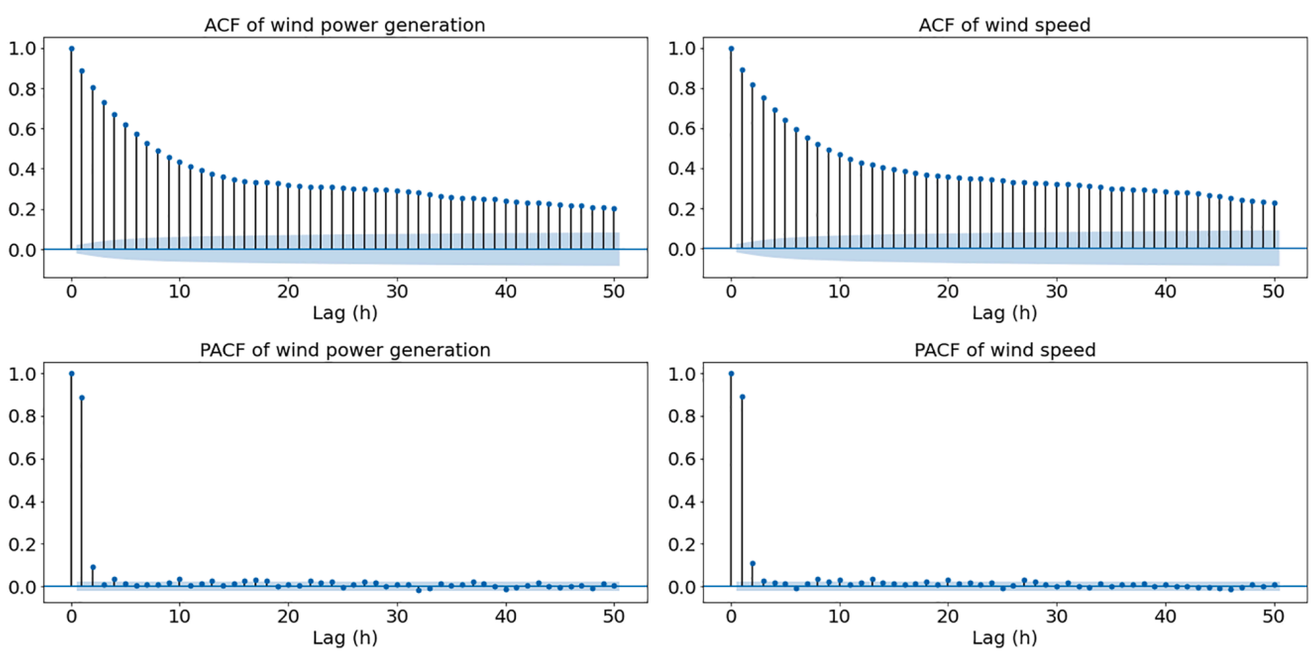


Fig. 5. ACF and PACF of the time series on wind power generation and wind speed in the year of 2020. The correlation outside the standard deviations are correlations and not a statistical fluke.

dependencies as the ACF show correlations outside the 95% confidence interval for all time lags (depicted as a blue area in Fig. 5). The values outside of the blue cone are very likely actual correlations. The PACF plots do not show any strong correlations except from the first three lags, and the correlations in the ACF plot after lag 3, are indirect correlations that can be explained by the first three time steps. Noteworthy, the ACF and PACF plots capture only linear dependencies in the wind power generation and wind speed, but there could be non-linear dependencies in the time series in addition. The deep learning models used in this work provide the capability to capture such non-linear relationships.

The correlation in the ACF and PACF plots shows that historical data on wind power generation should be used as input to the deep learning models to make predictions of the expected power generation. In addition, as wind power generation is directly affected by the amount of wind that hits the turbines, it is of interest to investigate which variables should be included (or excluded) as covariates to potentially improve the prediction performance. The exogenous variables included in addition to the historical data on wind power generation are the historical measurement data on weather conditions and the NWP from the AROME-Arctic model. The measured weather is included as input due to the complex terrain in the region of the wind farm, and the weather model with a spatial resolution of 2.5 km might be too coarse to capture local variations in the wind. Therefore, the information from measured wind from the nacelle could provide important additional information.

The prediction experiments are summarized as; 1) Use historical measurement data on power generation with historical measurement data on weather, and the 36 h ahead NWP as exogenous variables to predict the future power generation. 2) Use historical measurement data

on power generation, where only measured weather data is used as exogenous variables. 3) Use historical measurement data on power generation, where only the 36-h ahead NWP data is used as exogenous variables.

In Fig. 6 the example on the first prediction experiment where both measured weather (MW) and predicted weather (PW) is used as additional inputs to predict the future power generation is illustrated. In prediction experiments 2) or 3), either PW or MW is removed as input.

To make predictions with the different covariates, the DeepAR model proposed by [16] is selected as it has outperformed other models in several recent works on probabilistic forecasts with deep learning [19,16,32]. The inputs to the DeepAR models are the time series values (historical wind power generation) until $t-1$ and potential covariates (such as measured and predicted weather) at time t. The covariates and the time series values are thereafter concatenated before being fed into the internal units, which can either be LSTM or GRU layers. The output from the internal units is fed into two different linear layers. One for computing the mean, and one for computing the standard deviation. When computing the standard deviation, the linear layer is fed into a SoftPlus layer to ensure positive values. In the end, the computed mean and standard deviation are used as input to a Gaussian likelihood model where predictive samples are generated. During training, the model learns by maximizing a Gaussian log-likelihood function and is optimized via stochastic gradient descent with respect to the model parameters. The DeepAR model is described more in detail in Section 4 in [16].

In former works the default DeepAR configuration with internal LSTM units has been used [16,19,32]. In this work, the DeepAR model is tested with Gated Recurrent Units (GRU) [43] in addition to LSTM as internal units. A graphical illustration of the DeepAR model are presented in Fig. 7.

In this work, three models are selected to serve as benchmark models. The Auto-regressive Integrated Moving Average (ARIMA) model and two versions of a persistence model. The ARIMA (p,d,q) model can be used to define a large class of statistical models. The parameter p indicates the order of the auto-regressive component, d represents the initial differencing of the time series, and q the order of the moving average component. In addition, two versions of a persistence model are served as benchmarks for the DeepAR model. The persistence model assumes that the wind power generation at a certain future time will be the same as it is when the forecast is made (for instance, if the generation is 50 MW when the prediction is made, the persistence model assumes that the generation the next hour also is 50 MW). The persistence model has been a popular model to use as a benchmark as it is a simple method to implement, and are often a difficult method to outperform, especially in the range of 1-6 h ahead [21]. In addition, a modified version of the persistence model is served as a benchmark model in this work. The authors in [44] proposed a modified persistence model for predicting wind power generation. This model is a combination of the mean of the time series, where the future value is weighted as a function of the correlation between the current and average power, respectively. In prediction experiments with a forecast horizon above 3 h, the modified persistence model has been shown to outperform the original one [44].

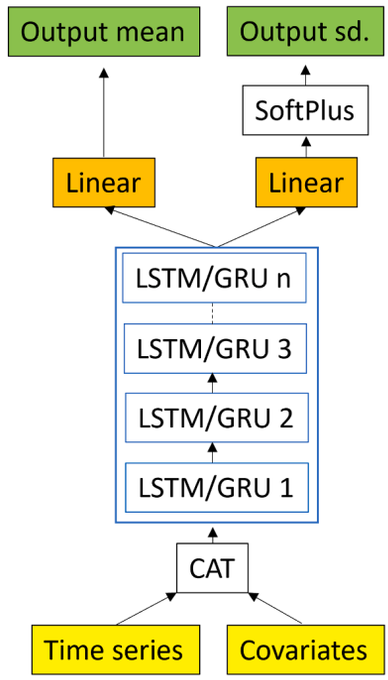


Fig. 7. Graphical illustration of the DeepAR structure. The time series values and the Covariates are concatenated (CAT) together before being fed into either the LSTM or GRU units. The block of internal units can vary between 1 and n layers depending on which configuration gives the highest prediction accuracy. The output is fed into different linear layers, one for computing the mean and one for the standard deviation (with a SoftPlus layer to ensure positive values). The output mean and standard deviation are used as input to a Gaussian likelihood model where the predictive samples are generated.

5.2. Prediction strategy

To train and evaluate the models, the time series is split into a training set (85%) and a test set (15%). The training set is further divided into training (80%) and validation (20%). The training set is used to fit the model parameters by minimizing the prediction loss, and the validation set is used to find the optimal configuration of the hyperparameters. Each model is set to train on 500 epochs, and to avoid potential overfitting during training, an early stopping rate is introduced.

5.3. Hyperparameter optimization with cross-validation

In this work, to find the optimal hyperparameter configuration for the DeepAR model a grid search in the hyperparameter space was performed. The searched hyperparameters include the length of the rolling window, which is the number of time points that the model gets to see before making the prediction, the number of hidden units, the number of layers, the dropout rate, and the learning rate. The search space for the selected hyperparameters and the optimum configuration for each

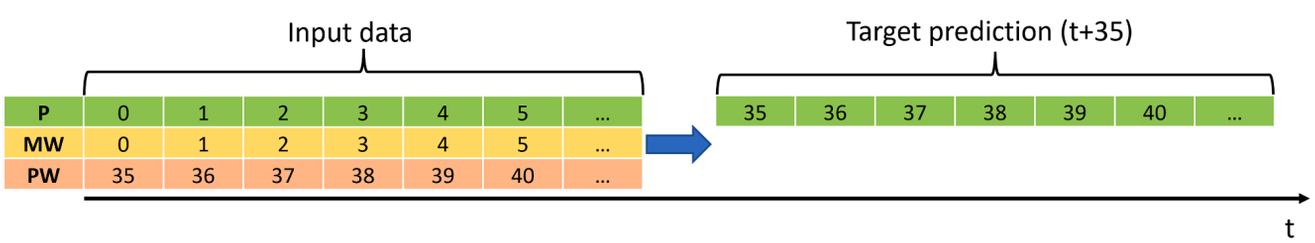


Fig. 6. The historical data on wind power generation (P) together with historical data on measured weather (MW) and predicted weather (PW) to predict the future wind power generation (target prediction).

model is detailed in Table 2. The configuration that resulted in the highest accuracy on the validation set is selected.

In order to determine the optimal order of the benchmark ARIMA model, a grid search of the different parameters was performed, and the model resulted in the best accuracy on the validation set in terms of the Akaike Information Criterion (AIC) was selected. For the time series of wind power generation, the ARIMA (2,0,1) configuration is the model providing the highest prediction performance.

6. Experimental evaluation and results

6.1. Evaluation metrics

When making point predictions, the main aim is to minimize the discrepancy between the predicted and true output, respectively. For the purpose of probabilistic forecasts, it is more complicated. Here one wants to minimize the loss between the predicted and true value for a specific quantile level (or percentile which are the upper and lower bound of the PI), but at the same time, one wants to have a PI that contains the true outcome. In addition, one does not want to have a PI that is too wide as it will contain no useful information, so the PI should be as sharp as possible, but still contain the true values. In the following, some popular metrics that will be used in the result section to provide the skill score of both point-and probabilistic forecasts are described.

A widely used metric to evaluate the performance of *point forecasts* is the root mean squared error (RMSE) and the Normalized RMSE (NRMSE). The RMSE is defined as

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\widehat{y}_i - y_i)^2}, \tag{2}$$

where $\widehat{y}$ and y are the predicted and true values. The RMSE measures the discrepancy between the predicted values and observed values at time i, over n number of observations. The NRMSE relates the RMSE to the observed average value in the observation period and is defined as:

$$NRMSE = \frac{RMSE}{\overline{y}}, \tag{3}$$

where $\overline{y}$ is the average of the observed values in the time series.

Another popular method to evaluate the accuracy of point forecasts are the mean absolute percentage error (MAPE). MAPE is defined as:

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \widehat{y}_i}{y_i}\right|, \tag{4}$$

The common feature of all these metrics (RMSE, NRMSE, and MAPE) is that the lower value, the higher accuracy.

The purpose of *probabilistic forecasts* is to have a PI that fulfills the calibration and sharpness criteria. In this section, some common metrics that are used to describe the performance of the PIs in terms of the sharpness and calibration criteria are presented.

The pinball-loss (PL), or the quantile-loss (QL) function is a common metric to measure the performance of the PI. For each quantile level, the PL function returns a value that can be interpreted as the accuracy of a quantile forecasting model. Let q be the target quantile, y the real value, and $\widehat{y}$ the quantile forecast, then the PL for quantile q can be written as:

$$PL_q\left(y,\widehat{y}\right) = \begin{cases} (y-\widehat{y})q & y \geqslant \widehat{y} \\ (\widehat{y}-y)(1-q) & \widehat{y} > y \end{cases}$$

The PL function penalizes the forecast if the model is over or under-predicting depending on the quantile level that is computed. The large quantile level will be more penalized for under-predicting than a low quantile level. Similarly, a low quantile level will be more penalized for over-predicting than a large quantile level. This makes sense as in the high quantile level case, one expects most of the observed values to be smaller than the predictions, and at the low quantile level one expects most of the values to be above the predicted values. Similar to the point forecast metrics, the lower PL, the more accurate the quantile forecast is. If computing the PL over a set of different quantiles levels, the final quantile loss result will be the average of all levels (often denoted as QLm).

In [16], the authors evaluated the sharpness, or the quantile risk for different quantile levels by considering the normalized sum, wQL of quantile losses. The wQL for a quantile q is computed as all pinball losses divided by the sum of true output:

$$wQL_q\left(y,\widehat{y}\right) = 2\frac{\sum_i PL_q\left(y_i,\widehat{y}_i\right)}{\sum_i y_i} \tag{5}$$

A low wQL indicate a sharper PI.

Besides sharpness, it is important to compute how calibrated the PI is. A popular calibration metric for probabilistic forecasts is the Prediction Interval Coverage Probability (PICP) [2]. The PICP is employed to compute the probability that the true outcome is within the PIs. The PCIP is defined as

$$PCIP = \frac{1}{n}\sum_{i=1}^{n} u_i, \tag{6}$$

where n is the total number of samples. When the true output is within the upper and lower bound, $u_i = 1$, otherwise $u_i = 0$. To obtain a well-calibrated PI, the coverage should be close as possible to the PI that is specified. For instance, for a 95% PI, the PCIP should be 0.95.

The final metric that is used to measure the quality of the PIs is the Mean Scaled Interval Score (MSIS). This metric was used as the preferred one in the M4 forecasting competition, where 100,000 time series and 61 forecasting methods were compared [45]. The MSIS score was proposed by [46] and evaluate the performances of the generated PIs as

$$MSIS = \frac{1}{h} \times \frac{\sum_{t=n+1}^{n+h}\left(U_t - L_t\right) + \frac{2}{\alpha}\left(L_t - Y_t\right)\mathbb{1}\mathbb{Y}_t < \mathbb{L}_t + \frac{2}{\alpha}\left(\mathbb{Y}_t - \mathbb{U}_t\right)\mathbb{1}\mathbb{Y}_t > \mathbb{U}_t}{\frac{1}{n-m}\sum_{t=m+1}^{n}\left|Y - Y_{t-m}\right|}, \tag{7}$$

where the lower and upper bounds of the PI are denoted by L_t and U_t, respectively. Y_t is the future observed values, h is the forecast horizon, and $\mathbb{1}$ is the indicator function (1 if Y_t is within the PI and 0 otherwise). Here α is the significance level, and for a 95% PI, α is set to 0.05. The MSIS metric deals with the sharpness and calibration criteria. It both penalize wide PIs (since U_t and L_t will be large), and penalize non-coverage. Here, a lower MSIS score indicates a better PI in terms of sharpness and calibration.

Table 2
Details of the hyperparameter search space and optimal configurations for different models.

Parameter	Search space	Optimal $DeepAR_{GRU}$	Optimal $DeepAR_{LSTM}$
Rolling window length	36, 72	36	36
Layers	1, 2, 3	2	2
Hidden units	32, 64	32	64
Dropout rate	0.0, 0.1, 0.2	0.0	0.2
Learning rate	10^{-4},10^{-3},10^{-2}	10^{-2}	10^{-3}

6.2. Results and discussion

In Table 3 the results are given. In this work, the 95% PI is computed as it is a widely used choice for economic, financial, and energy-related

Table 3
36-h ahead prediction scores with different dataset configurations.

Model	NRMSE	MAPE	MSIS	$PCIP_{2.5}$	$PCIP_{97.5}$	Mean wQL
Configuration 1: Measured and predicted weather (wind speed + wind direction)						
$DeepAR_{GRU}$	0.17	0.16	3.56	0.00	1.00	0.028
$DeepAR_{LSTM}$	**0.16**	**0.15**	**2.53**	**0.027**	**0.972**	**0.020**
Configuration 2: Measured weather (wind speed + wind direction)						
$DeepAR_{GRU}$	0.45	0.41	10.73	0.00	0.55	0.084
$DeepAR_{LSTM}$	0.40	0.35	10.16	0.00	0.55	0.080
Configuration 3: Predicted weather (wind speed + wind direction)						
$DeepAR_{GRU}$	0.24	0.21	5.32	0.00	1.00	0.042
$DeepAR_{LSTM}$	0.29	0.25	4.03	0.00	**0.972**	0.032
Benchmark models						
ARIMA	0.52	0.44	4.34	0.00	1.00	0.037
Persistence	1.06	6.17	–	–	–	–
Modified Persistence	0.79	5.09	–	–	–	–

forecasting applications [45]. The point prediction scores for predicting the 36-h ahead power generation is given in terms of NRMSE and MAPE (Eq. (3) and (4)). For probabilistic forecasts, the scores are given in terms of MSIS, PCIP, and the wQL (Eq. (5)–(7). In Table 3, the mean wQL for quantile level 0.025 and 0.975 is computed.

Table 3 shows that the DeepAR model with LSTM units resulted in the best accuracy for all experiments. As a downside, the training time for this model was significantly larger than the DeepAR model with GRU units. The training time per epoch was approximately 17–18 s and 39–40 s with GRU and LSTM units, respectively. To obtain the optimized models, the models were trained for approximately 230 epochs. This results in a total training time of 65 and 150 min with GRU and LSTM units, respectively. The models were trained on a NIVIDA Tesla K80 hardware. Among the benchmark models, the ARIMA model obtained the best prediction performance, but is less accurate than the DeepAR model. The Persistence models obtained the worst prediction performance for all experiments. The probabilistic scores are not reported for the persistence models as they do not provide the capability to make probabilistic forecasts.

The best performance was obtained with the dataset configuration where both measured and predicted weather was included as exogenous variables. With this configuration, the $DeepAR_{LSTM}$ model computed the sharpest PI as the mean wQL and the MSIS are low. In addition, this model computed a PI that is well-calibrated as the PCIP is close to the quantile levels that are specified. For instance, for $PCIP_{2.5}$, the computed PI resulted in a PCIP of 2.7%, which indicates that it is a probability of 2.7% that the true outcome is below this quantile level. The PCIP for the upper boundary is also close to the specified quantile level.

Using only measured weather as exogenous variables resulted in the worst prediction performance for both models. The computed PI for this configuration is neither sharp nor calibrated, as the MSIS and wQL are large and the $PCIP_{97.5}$ value is 0.55. This shows that the true outcome is below the upper PI boundaries only 55% of the time, which is not adequate in a 95% PI interval.

When the NWP is used as the only exogenous variable, acceptable prediction performances were obtained for the day-ahead predictions. However, indicated by the MSIS and wQL scores, the PI is less sharp in this case. In addition, the PI is not fully calibrated for both models as there is zero probability that the true outcome is below the lower PI boundary.

The results indicate that combining historical measurement data and NWPs helps improve the day-ahead prediction of wind power generation. This shows that adding historical data on measured weather allows the DeepAR model to auto-correct systematic biases in the NWPs.

In Fig. 8 the 36-h ahead predictions with the best $DeepAR_{LSTM}$ model in terms of calibration and sharpness is provided (configuration 1 with measurements and NWPs combined). In addition, the $DeepAR_{LSTM}$ for configuration 2 and 3 is shown. These are two examples of PI results that are miscalibrated and have low sharpness. In the following illustration, the 50% PI and the median prediction are included.

The red line represents observed values, while the green line represents the median value (or point forecast) of the probabilistic forecast. It is seldom a perfect match between the predicted point forecast and the actual value. This is not surprising as making accurate point forecasts of the day-ahead wind power generation is a very difficult problem. The green nuances in the graphs represent the different PIs. The uppermost graph in Fig. 8 show the resulting prediction with the $DeepAR_{LSTM}$ model using configuration 1. In this prediction, the 95% PI indicated by the bright green color shows that the actual measurements fall within the interval approximately 95% of the time and therefore show a well-calibrated PI. The observed values fall outside the interval at some points in the period between 12:00 and 18:00 the 30-Dec. This is acceptable, as given by the 95% requirement, some values may fall outside the range. The rest of the time the observed values are within the 95% PI. The graph also shows that the model is more confident at the beginning of the prediction period as the PI boundaries are very sharp, but the uncertainties increase as longer ahead in the future the predictions are made.

On contrary, the graph in the middle shows the $DeepAR_{LSTM}$ model for the configuration using only measured wind. This shows a low degree of sharpness and is not calibrated as several observations are outside the PI boundaries. The lowermost graph shows the result from the $DeepAR_{LSTM}$ model where only the predicted wind is used as exogenous variables. This result shows a well-calibrated PI as most observations are within the boundaries, but the PI is less informative as is it wide and consequently provides less useful information about the variations in output.

Due to technical limitations, the wind power plant can maximum produce 54 MW (18 turbines $\times$ 3 MW). However, all models in Fig. 8 compute a PI that has upper boundaries that exceed this value. This is not possible due to the theoretical maximum of 54 MW, and one can be sure that there is zero probability that the total wind power generation will exceed this value. Therefore, all values above 54 MW in the upper boundary of the computed 95% PIs can be replaced by the theoretical maximum power generation. In Fig. 9 the modified 95% PI of the $DeepAR_{LSTM}$ with the best prediction performance is provided. Here all values above the theoretical maximum are replaced by the theoretical maximum of 54 MW.

Fig. 9 show that replacing the upper PI boundaries with the theoretical maximum power output from the wind farm generates a PI that is much sharper and is very accurate at several time steps. This indicates that when making a probabilistic forecast of power output from wind farms, knowledge regarding technical limitations could contribute to achieving even more accurate PIs. The scores of the original and adjusted 95% PI for the $DeepAR_{LSTM}$ for configuration 1 is given in Table 4.

From Table 4 the MSIS and mean wQL show that the adjusted PI is sharper than the original PI. The PCIP for the upper and lower boundaries are the same for both PIs and are not reported.

7. Conclusions

In this work, we investigated the performance of deep learning methods for probabilistic forecasting of wind power production, which is characterized by a highly intermittent nature. We focused on one day-ahead forecast with the DeepAR model, which has achieved state-of-the-art performances in several time series forecasting tasks.

Our experiments focused on evaluating which covariates that are useful for the prediction performance and we investigated the importance of including historical measurements of wind and power

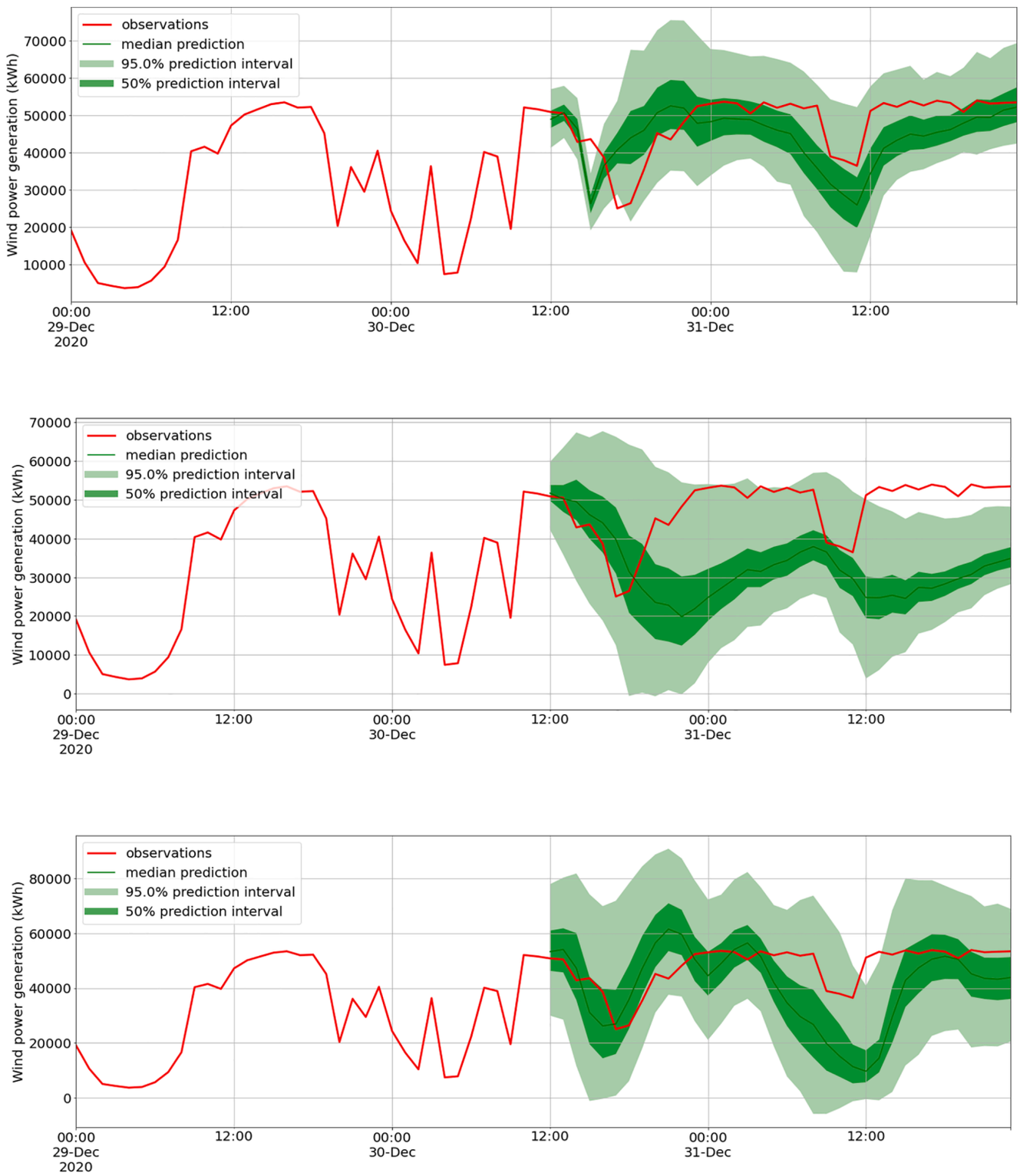


Fig. 8. 36 h ahead predictions of wind power generation. The uppermost Figure show the DeepAR$_{LSTM}$ model for configuration 1, the Figure in the middle shows the DeepAR$_{LSTM}$ model for configuration 2, and the lowermost Figure shows the result from the DeepAR$_{LSTM}$ model for configuration 3.

generation. We tested two different Recurrent Neural Network layers in the DeepAR model, namely LSTM and GRU, and evaluated the quality of the PI generated by the model. We also compared against ARIMA and two baseline models, commonly used to perform wind power forecasting, which do not leverage historical information. The DeepAR model with LSTM units obtained the most accurate prediction performance for all experiments and outperformed the baseline models.

Among the different configurations, the best performance in terms of sharpness and calibration was obtained when both historical data on measured weather and the NWPs were used as exogenous variables, resulting in an MSIS of 2.53. When using only the predicted weather as exogenous variables, worse results were obtained with an MSIS of 4.03 and 10.16, respectively. This could be due to the highly complex topography of the region where the wind farm is located, which makes it increasingly difficult to make accurate weather forecasts. We hypothesize that using the historical measured wind allows the deep learning

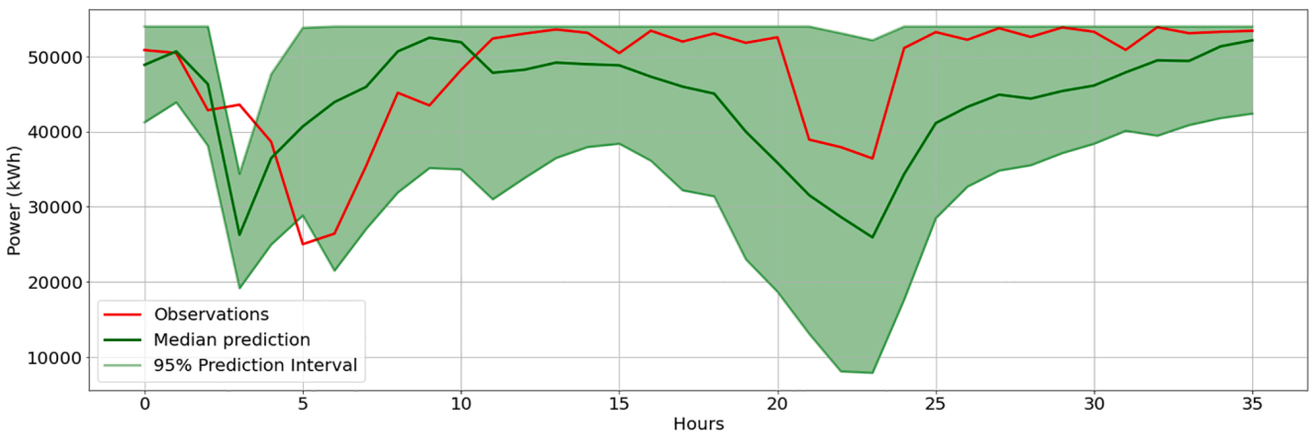


Fig. 9. 36 h ahead probabilistic forecast where the upper boundaries of PI are adjusted with respect to the theoretical maximum power generation of 54 MW.

Table 4
Probabilistic performance scores with the original and adjusted 95% PI.

	Original 95% PI		Adjusted 95% PI	
Model	MSIS	mean wQL	MSIS	mean wQL
$DeepAR_{LSTM}$	2.53	0.020	1.16	0.013

model to correct systematic biases, which are common in the NWP of certain regions as RES technologies are directly affected by the current weather. Using historical weather measurements as input allows the prediction model to compensate for the errors in weather predictions, which is something that could improve the forecast of power generation. In addition, knowledge of the physical properties and the technical limitations of the wind power plant is extremely important to calibrate and correct the output of a machine learning model. Remarkably, we adjusted the upper limit of the PI, based on the maximum power output of a turbine, and obtained a much sharper PI.

The results from our study can serve as a reference for both the research community and industry, as it shows the importance of carefully selecting the factors to be considered when training a machine learning model for energy analytics applications.

CRediT authorship contribution statement

Odin Foldvik Eikeland: Conceptualization, Methodology, Software, Investigation, Writing - original draft. **Finn Dag Hovem:** Resources. **Tom Eirik Olsen:** Resources. **Matteo Chiesa:** Conceptualization, Supervision. **Filippo Maria Bianchi:** Conceptualization, Methodology, Validation, Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

O.F.E, M.C, and F.M.B acknowledge the support from the research project “Transformation to a Renewable & Smart Rural Power System Community (RENEW)”, connected to the Arctic Centre for Sustainable Energy (ARC) at UiT-the Arctic University of Norway through Grant No. 310026.

References

[1] Liu Y, Qin H, Zhang Z, Pei S, Jiang Z, Feng Z, Zhou J. Probabilistic spatiotemporal wind speed forecasting based on a variational bayesian deep learning model. Appl Energy 2020;260:114259.
[2] Huang Q, Wei S. Improved quantile convolutional neural network with two-stage training for daily-ahead probabilistic forecasting of photovoltaic power. Energy Convers Manage 2020;220:113085.
[3] Ruhnau O, Hennig P, Madlener R. Economic implications of forecasting electricity generation from variable renewable energy sources. Renew Energy 2020;161:1318–27.
[4] Mazzi N, Pinson P. Wind power in electricity markets and the value of forecasting. In: Renewable Energy Forecasting. Elsevier; 2017. p. 259–78.
[5] Ahmed Mohammed A, Aung Z. Ensemble learning approach for probabilistic forecasting of solar power generation. Energies 2016;9(12):1017.
[6] Brusaferri A, Matteucci M, Portolani P, Vitali A. Bayesian deep learning based method for probabilistic forecast of day-ahead electricity prices. Appl Energy 2019;250:1158–75.
[7] Zhang H, Liu Y, Yan J, Han S, Li L, Long Q. Improved deep mixture density network for regional wind power probabilistic forecasting. IEEE Trans Power Syst 2020;35(4):2549–60.
[8] Toubeau J-F, Bottieau J, Vallée F, De Grève Z. Deep learning-based multivariate probabilistic forecasting for short-term scheduling in power markets. IEEE Trans Power Syst 2018;34(2):1203–15.
[9] Wang H, Yi H, Peng J, Wang G, Liu Y, Jiang H, Liu W. Deterministic and probabilistic forecasting of photovoltaic power based on deep convolutional neural network. Energy Convers Manag 2017;153:409–22.
[10] Wang H-Z, Li G-Q, Wang G-B, Peng J-C, Jiang H, Liu Y-T. Deep learning based ensemble approach for probabilistic wind power forecasting. Appl Energy 2017;188:56–70.
[11] Jin H, Shi L, Chen X, Qian B, Yang B, Jin H. Probabilistic wind power forecasting using selective ensemble of finite mixture gaussian process regression models. Renew Energy 2021;174:1–18.
[12] Gougheri SS, Jahangir H, Golkar MA, Ahmadian A, Golkar MA. Optimal participation of a virtual power plant in electricity market considering renewable energy: A deep learning-based approach. Sustain Energy, Grids Netw 2021;26:100448.
[13] Sadeghi S, Jahangir H, Vatandoust B, Golkar MA, Ahmadian A, Elkamel A. Optimal bidding strategy of a virtual power plant in day-ahead energy and frequency regulation markets: A deep learning-based approach. Int J Electr Power Energy Syst 2021;127:106646.
[14] Jahangir H, Gougheri SS, Vatandoust B, Golkar MA, Ahmadian A, Hajizadeh A. Plug-in electric vehicle behavior modeling in energy market: A novel deep learning-based approach with clustering technique. IEEE Trans Smart Grid 2020;11(6):4738–48.
[15] Xiang L, Li J, Hu A, Zhang Y. Deterministic and probabilistic multi-step forecasting for short-term wind speed based on secondary decomposition and a deep learning method. Energy Convers Manage 2020;220:113098.
[16] Salinas D, Flunkert V, Gasthaus J, Januschowski T. Deepar: probabilistic forecasting with autoregressive recurrent networks. Int J Forecast 2020;36(3):1181–91.
[17] Zhang W, Quan H, Srinivasan D. An improved quantile regression neural network for probabilistic load forecasting. IEEE Trans Smart Grid 2018;10(4):4425–34.
[18] Afrasiabi M, Mohammadi M, Rastegar M, Afrasiabi S. Advanced deep learning approach for probabilistic wind speed forecasting. IEEE Trans Industr Inf 2020;17(1):720–7.
[19] Mashlakov A, Kuronen T, Lensu L, Kaarna A, Honkapuro S. Assessing the performance of deep learning models for multivariate probabilistic energy forecasting. Appl Energy 2021;285:116405.
[20] Zhang Y, Wang J, Wang X. Review on probabilistic forecasting of wind power generation. Renew Sustain Energy Rev 2014;32:255–70.

[21] Monteiro C, Bessa R, Miranda V, Botterud A, Wang J, Conzelmann G, et al. Wind power forecasting: State-of-the-art 2009, Tech. rep., Argonne National Lab. (ANL), Argonne, IL (United States) (2009).
[22] Zhang Y, Li Y, Zhang G. Short-term wind power forecasting approach based on seq2seq model using nwp data. Energy 2020;213:118371.
[23] Higashiyama K, Fujimoto Y, Hayashi Y. Feature extraction of nwp data for wind power forecasting using 3d-convolutional neural networks. Energy Procedia 2018; 155:350–8.
[24] Wang X, Guo P, Huang X. A review of wind power forecasting models. Energy Procedia 2011;12:770–8.
[25] Kavasseri RG, Seetharaman K. Day-ahead wind speed forecasting using f-arima models. Renew Energy 2009;34(5):1388–93.
[26] Zhao W, Wei Y-M, Su Z. One day ahead wind speed forecasting: a resampling-based approach. Appl Energy 2016;178:886–901.
[27] Wang J, Song Y, Liu F, Hou R. Analysis and application of forecasting models in wind power integration: a review of multi-step-ahead wind speed forecasting models. Renew Sustain Energy Rev 2016;60:960–81.
[28] Hong T, Fan S. Probabilistic electric load forecasting: a tutorial review. Int J Forecast 2016;32(3):914–38.
[29] Chen Y, Kang Y, Chen Y, Wang Z. Probabilistic forecasting with temporal convolutional neural network. Neurocomputing 2020;399:491–501.
[30] Gasthaus J, Benidis K, Wang Y, Rangapuram SS, Salinas D, Flunkert V, Januschowski T. Probabilistic forecasting with spline quantile function rnns. In: in: The 22nd international conference on artificial intelligence and statistics, PMLR; 2019. p. 1901–10.
[31] Nguyen N, Quanz B. Temporal latent auto-encoder: a method for probabilistic multivariate time series forecasting. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35; 2021. p. 9117–25.
[32] Alexandrov A, Benidis K, Bohlke-Schneider M, Flunkert V, Gasthaus J, Januschowski T, Maddix DC, Rangapuram SS, Salinas D, Schulz J, et al. Gluonts: probabilistic and neural time series modeling in python. J Mach Learn Res 2020;21 (116):1–6.
[33] Jahangir H, Golkar MA, Alhameli F, Mazouz A, Ahmadian A, Elkamel A. Short-term wind speed forecasting framework based on stacked denoising auto-encoders with rough ann. Sustain Energy Technol Assess 2020;38:100601.
[34] Wang J, Wang S, Li Z. Wind speed deterministic forecasting and probabilistic interval forecasting approach based on deep learning, modified tunicate swarm algorithm, and quantile regression. Renew Energy 2021;179:1246–61.
[35] Liu H, Duan Z, Chen C, Wu H. A novel two-stage deep learning wind speed forecasting method with adaptive multiple error corrections and bivariate dirichlet process mixture model. Energy Convers Manage 2019;199:111975.
[36] Zhao X, Liu J, Yu D, Chang J. One-day-ahead probabilistic wind speed forecast based on optimized numerical weather prediction data. Energy Convers Manage 2018;164:560–9.
[37] Zhu S, Yuan X, Xu Z, Luo X, Zhang H. Gaussian mixture model coupled recurrent neural networks for wind speed interval forecast. Energy Convers Manage 2019; 198:111772.
[38] Jahangir H, Tayarani H, Gougheri SS, Golkar MA, Ahmadian A, Elkamel A. Deep learning-based forecasting approach in smart grids with microclustering and bidirectional lstm network. IEEE Trans Industr Electron 2020;68(9):8298–309.
[39] Al-Gabalawy M, Hosny NS, Adly AR. Probabilistic forecasting for energy time series considering uncertainties based on deep learning algorithms. Electr Power Syst Res 2021;196:107216.
[40] Kuleshov V, Fenner N, Ermon S. Accurate uncertainties for deep learning using calibrated regression. In: International Conference on Machine Learning. PMLR; 2018. p. 2796–804.
[41] La Salle JLG, Badosa J, David M, Pinson P, Lauret P. Added-value of ensemble prediction system on the quality of solar irradiance probabilistic forecasts. Renew Energy 2020;162:1321–39.
[42] Sun M, Feng C, Zhang J. Multi-distribution ensemble probabilistic wind power forecasting. Renew Energy 2020;148:135–49.
[43] Bianchi FM, Maiorino E, Kampffmeyer MC, Rizzi A, Jenssen R. Recurrent neural networks for short-term load forecasting: an overview and comparative analysis; 2017.
[44] Moehrlen C. Uncertainty in wind energy forecasting, Ph.D. thesis, University College Cork; 2004.
[45] Makridakis S, Spiliotis E, Assimakopoulos V. The m4 competition: 100,000 time series and 61 forecasting methods. Int J Forecast 2020;36(1):54–74.
[46] Gneiting T, Raftery AE. Strictly proper scoring rules, prediction, and estimation. J Amer Statist Assoc 2007;102(477):359–78.

9

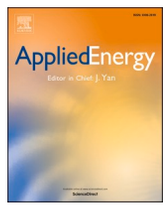
Paper IV

Applied Energy 348 (2023) 121572

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy



Power availability of PV plus thermal batteries in real-world electric power grids

Odin Foldvik Eikeland [a,b], Colin C. Kelsall [a], Kyle Buznitsky [a], Shomik Verma [a], Filippo Maria Bianchi [c], Matteo Chiesa [a,b,d,*], Asegun Henry [a,**]

[a] Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
[b] Department of Physics and Technology, UiT – The Arctic University of Norway, 9037 Tromsø, Norway
[c] Department of Mathematics and Statistics, UiT – The Arctic University of Norway Technology, 9037 Tromsø, Norway
[d] Laboratory for Energy and NanoScience (LENS), Khalifa University of Science and Technology, Masdar Institute Campus, 127788 Abu Dhabi, United Arab Emirates

HIGHLIGHTS

- The power availability from photovoltaics and thermal battery was investigated.
- Novel thermal battery technologies can improve renewable energy dispatchability.
- The power availability vastly improves if CO_2 emissions are reduced.

ARTICLE INFO

Keywords:
Thermal energy grid storage
Electric power system modeling
Electric power system decarbonization
Power availability

ABSTRACT

As variable renewable energy sources comprise a growing share of total electricity generation, energy storage technologies are becoming increasingly critical for balancing energy generation and demand.

In this study, a real-world electricity system was modeled rather than modeling hypothetical future electric power systems where the existing electricity infrastructure are neglected. In addition, instead of modeling the general requirements of storage in terms of cost and performance, an existing thermal energy storage concept with estimated capital cost that are sufficiently low to enable large-scale deployment in the electric power system were modeled. The storage unit is coupled with a photovoltaic (PV) system and were modeled with different storage capacities, whereas each storage unit had various discharge capacities.

The modeling was performed under a baseline case with no emission constraints and under hypothetical scenarios in which CO_2 emissions were reduced. The results show that power availability increases with increasing storage size and vastly increases in the hypothetical CO_2 reduction scenarios, as the storage unit is utilized differently. When CO_2 emissions are reduced, the power system must be less dependent on fossil fuel technologies that currently serve the grid, and thus rely more on the power that is served from the PV + storage unit.

The proposed approach can provide increased knowledge to power system planners regarding how adding PV + storage systems to existing grids can contribute to the efficient stepwise decarbonization of electric power systems.

1. Introduction

The use of variable renewable energy (VRE) resources, such as wind power and solar photovoltaics (PV), is expanding rapidly as a share of total power generation and is critical to the decarbonization of electrical power systems [1–3]. The weather-dependent intermittency of VRE sources complicates the planning and management of power systems as the electric power generation can no longer be directly modulated to match the electricity demand. Energy storage will therefore be an increasingly critical component of future energy systems with high

* Corresponding author at: Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA.
** Corresponding author.
E-mail addresses: mchiesa@mit.edu (M. Chiesa), ase@mit.edu (A. Henry).

https://doi.org/10.1016/j.apenergy.2023.121572
Received 30 January 2023; Received in revised form 26 June 2023; Accepted 8 July 2023
Available online 25 July 2023

penetrations of VRE sources. Energy storage can charge excess electricity in periods with high generation and low demand, and then discharge the electricity in periods with low generation from the VRE sources to match the load in periods with high electricity demand [4,5]. In addition, energy storage stabilizes the grid by providing additional electricity supply when there is a surge in electricity demand or a sudden drop in supply from the VRE sources. Energy storage also enables cost reduction of the grid by allowing for an increased share of cheap VRE technologies in the electricity supply portfolio [6], and reduces potential curtailment of the electricity generation from VRE sources during periods when the generation exceeds the demand [7].

The need for inexpensive storage over periods with different lengths, from seconds to days and even seasonal storage, has accelerated in accordance with the increasing share of VRE technologies in electric power systems [6,8]. Pumped hydropower storage (PHS) and compressed air energy storage (CAES) are well-established technologies for large-scale energy storage, although they are only applicable in a few geographic areas. Hydrogen storage is thought to be a promising long-term storage solution. However, due to the high capital cost of charging and discharging, it has the greatest potential in the seasonal storage regime for getting sufficient low energy storage capital cost [9].

Lithium-ion batteries have been the state-of-the-art technology for short-term storage. However, capital costs between US$80 and US$100 kWh^{-1} make them unaffordable for the multi-day storage objectives required to completely decarbonize the grid [4,5,10]. Concentrated solar power with thermal energy storage (CSP-TES) has been seen as a promising option, but major projects around the world have been plagued by delays, cost overruns and mechanical issues, and interest has waned in recent years [11,12]. Studies suggest that achieving cost-efficient multi-day storage requires a capital cost reduction to US $3–30 kWh^{-1} [5,13]. Resolving this issue could enable more rapid decarbonization of the power system, resulting in a 25% reduction in global GHG emissions [14,15]. Therefore, one of the most significant issues that needs to be resolved to achieve the GHG emission reduction targets is to enable cost-effective pathways for increasingly implementing energy storage technologies into the electricity system.

A storage concept based on Thermal Energy Storage (TES) has shown promising potential to achieve sufficiently low capital cost in the multi-day storage regime. TES stores the electricity as heat rather than electrochemically, and then converts it back to electricity when needed [16]. The Thermal Energy Grid Storage (TEGS) concept, detailed in [17], stores electricity as sensible heat in graphite storage blocks and uses thermophotovoltaics (TPV) to convert heat back to electricity on demand [17,18]. While the conversion of heat to electricity results in significant efficiency penalties, storing energy as heat instead of electrochemically can be vastly cheaper, and thus the round-trip efficiency (RTE) penalty compared to electrochemical batteries ($\sim$ 90%) can potentially be a worthwhile tradeoff [17]. To maximize the conversion efficiency from heat to electricity, the heat is stored at extremely high temperatures ($\sim$2400 °C). In a recent work by [19], the authors demonstrated a world-record high conversion efficiency of 41% using TPV, and reported a projected conversion efficiency of 50% in the future. As such this technology can achieve a projected cost below US$ 20 kWh^{-1} at gigawatt scales.

In addition to the projected low cost, a unique property of TEGS compared to Li-ion battery technology is the fact that, since energy is stored as heat in graphite blocks and thereafter converted to electricity using TPV, it enables the possibility of fully decoupling the charge and discharge capacities of the storage unit. This allows the TEGS to charge (i.e., store heat) at a much higher capacity than that required for discharging. The benefit of such a mechanism is that a large amount of energy can be charged in a short amount of time when generation surpluses exist and discharged over a longer period to cover the electricity load in periods where demand exceeds supply. The TEGS system also has advantages in terms of durability, safety, and replaceability which make this technology a promising option to adopt into decarbonized electricity systems. In comparison to the state-of-the art Li-ion batteries, the TEGS system is more durable due to the construction materials. While the lifetime of electrochemical Li-ion batteries is affected by the depth of discharge and the number of cycles, the construction materials (graphite and tin) have no clear degradation mechanism and enables TEGS system to have an expected lifetime of 30 years or more (while li-ion are replaced after approximately 10 years). All construction materials of the TEGS system are at thermodynamic equilibrium giving no risk for chemical reactions. Additionally, it is housed in an inert environment with no immediate access to oxygen, preventing fire hazard. In terms of replaceability, the TEGS unit consists of separate components (Graphite storage blocks for storing heat, and a power block with thermal photovoltaics for generating electricity) which can be replaced separately if needed.

The body of existing literature counts several studies that have employed different approaches to evaluate the value of using storage to increase the dispatchability of VRE sources, and the different studies have highlighted the storage requirements (capital cost and storage duration) to enable the full decarbonization of the power system. Table 1 shows an overview of some relevant studies, where the key findings in each work are highlighted.

The vast number of previous studies on modeling the value of energy storage in emerging power systems, mainly focus on modeling hypothetical future electric power systems starting from scratch (i.e., “greenfield” models) [6,20–25]. However, such studies can, in many

Table 1
Overview of relevant work addressing the value of energy storage.

Ref.	Year	Key findings
[21]	2016	Large-scale deployment of available battery technologies requires cost reductions
[24]	2016	Pathways to fully renewable systems are feasible with high cost and overgeneration.
[4]	2016	Cost reduction for storage technologies is required to reach widespread profitability.
[22]	2017	The availability of how low-carbon technologies impact the optimal capacity mix and generation patterns were demonstrated.
[8]	2018	To reliably meet 100% of total annual electricity demand, weeks of energy storage are required to support with electricity.
[20]	2018	The role of energy storage units in power systems with high shares of VRE was analyzed. The importance of storage increases with the increasing share of renewable-based power technologies.
[23]	2018	Firm low-carbon resources consistently lower decarbonized electricity system costs, and the availability of firm low-carbon resources reduces costs 10%–62% in zero-CO_2 cases
[25]	2019	The benefits of hydro power and storage units were analyzed. Three decarbonized power systems with distinct grid expansion strategies were compared. Cutting transmission volume does not increase the total costs.
[7]	2019	Curtailment of renewable energy generation can be avoided using energy storage.
[5]	2019	Energy storage cost below $20 kWh^{-1} can enable cost-competitive baseload power.
[30]	2020	Hydrogen storage with up to 2 weeks of discharge duration is expected to be cost-effective in future power systems.
[31]	2020	Hydrogen storage enable for sector coupling in real-world power systems.
[32]	2020	Electricity triangle assures a consistent framework for the energy transition.
[10]	2020	Current Li-ion capital cost exceed storage value in many instances.
[26]	2020	Decarbonization is less expensive with Energy Storage Systems, given sufficient low-cost assumptions
[6]	2021	Energy capacity costs must be $\leq$US$20/kWh to reduce the electricity price by $\geq$10%.
[33]	2021	Power systems with 100% RE is possible using existing technologies.
[34]	2021	There is a need for analytic tool development to model how to achieve a power system that are 100% decarbonized.
[35]	2021	Clean firm resources are cost-effective in decarbonizing the grid
[9]	2022	Green hydrogen cost between $0.79/kg and $1.94/kg in 2030 can be achieved
[27]	2022	The demand for power capacity will drive future adoption of higher battery power capacity

cases, lead to vague results as this requires a complete change of the current electricity mix. This is in many cases challenging to implement due to policy considerations. In addition, by modeling greenfield cases, infrastructure that already exists is neglected.

Another approach to modeling the electric power system is by studying scenarios using the current electricity infrastructure (i.e., "brownfield studies). The benefit of such studies compared to greenfield studies is that this enables insight into how the existing electricity system can transform towards decarbonization from the current electricity infrastructure and hence give insight into which measures that must be taken to decarbonize the current electricity system.

Some former literature has applied a brownfield modeling approach when studying decarbonization pathways but lacks in modeling the potential value of using specific storage technologies [5,10,26–28]. These studies model the storage requirements in general, whereas all studies show that the capital cost must be below US \$20 kWh^{-1}. No previous work has modeled the potential of using an emerging storage concept based on TES that already exists on a lab scale [17,18].

In this study, a framework for addressing the value of using TEGS that has sufficiently low capital cost to be economically used in an electric grid is proposed. Using a Capacity Expansion Model (CEM) [29], a hypothetical PV + TEGS system that is interconnected to an existing real-world grid in the Northeastern US is considered. The TEGS unit charges excess electricity from PV during periods of surplus generation. When the grid demands electricity and the PV plant cannot deliver sufficient power due to a lack of solar availability, the stored energy is discharged. Different storage sizes with varying discharge capacities connected to the PV plant are modeled to optimize power availability. To investigate how emission constraints affect the energy availability of PV + storage systems, a hypothetical future scenario is modeled for the existing power system where CO_2 emissions are reduced.

The main contributions of this study are: Rather than modeling the power system as a greenfield case study, an abstract representation of an existing grid, i.e., a "brownfield" model is analyzed to address how adding a PV + storage system can contribute to decarbonizing the grid. Instead of modeling general requirements of storage to enable the full decarbonization of the power system, a TES unit that currently exists at lab-scale and has promising cost projections that are well-documented in the literature is modeled [17]. This study can provide increased knowledge to power system planners on how coupling emerging storage technologies and PV systems to existing grids can contribute to stepwise decarbonizing of the grid in a more short-term horizon.

2. Methods

2.1. Modeled electric power system

In this study, an idealized single node representing the electric grid region in the New England power system in North America is modeled. This system considers one grid zone that represents a simplified power system topology of the states of Massachusetts, New Hampshire, and Rhode Island. The electricity demand, capital cost and performance data for the different generation technologies in these regions were collected from the NREL annual technology baseline (ATB) and the U.S. Energy Information Administration (EIA). The Github library PowerGenome[1] was used to collect the input data and shape them to the required format for the CEM. The weather data used to construct the hourly generation profiles for solar and wind resources are collected from Vibrant Clean Energy (without any usage restrictions) using PowerGenome. The data from Vibrant Clean Energy is obtained from the National Oceanic and Atmospheric Administration (NOAA) High-Resolution Rapid Refresh (HRRR) weather forecast model. The weather forecast model is run every hour over a 3-km horizontal resolution that covers the United States. The weather year for modeling VRE availability in this study was the year of 2020.

The average annualized electricity demand for the modeled grid is 9.4 GWh, with a peak load of 16.7 GWh. Fig. 1 illustrates the hourly electricity demand in the modeled power system.

The blue graph shows the electricity demand with an hourly resolution, while the black graph shows the running average electricity demand with a weekly resolution. Clearly, the electricity demand is highest during summer due to increased usage of air conditions in high-temperature periods, and the lowest electricity usage is during the spring and autumn period when there is little need for heating and cooling.

At the supply side, the total installed generation capacity for the modeled grid zone is 15 gigawatts (GW). Fig. 2 shows the share of installed capacity for the different technologies. Natural Gas (NG) is the dominant power supply technology, accounting for 59% of the installed capacity. In the existing power grid, VRE sources such as wind and solar PV represent a smaller share (14%) of the overall electricity generation mix.

The hypothetical PV + storage power system is connected to the existing power system through a transmission grid network. The transmission grid network has a maximum capacity of 200 MW. Fig. 2 shows how the hypothetical system is connected to the existing grid, where the combined system can fully participate in the power system by exchanging electricity on demand. In this study, the storage unit were chosen to be modeled in conjunction with a PV plant, which is believed to be the most dominant source of electricity generation in the future power market [1,2,36]. In addition, the normal profile of the daily generation from PV plants is believed to be a good match with storage technologies, as it can store electricity when the PV plant power generates a large amount of electricity during mid-day (and the demand is often low during mid-day) and discharge the stored electricity when the sun is set (during early morning and afternoon/evening).

In this study, PV plants with installed peak power capacities of 100 MW and 1 GW were analyzed (see the Supplementary material for the GW scale modeling), which is represents the range of typical sizes of utility-scale solar energy farms in the U.S. [37].The storage unit was modeled with different energy storage capacities and are specified in the storage modeling Section 2.2.2. The modeled system does not influence the overall electricity price in the grid and is therefore considered a price taker.

2.1.1. Thermal energy grid storage (TEGS)

To charge the TEGS unit, excess electricity is used to fuel resistive heating materials (graphite), transforming the electricity into heat at a temperature exceeding 2500 degrees Celsius. Then, the energy is transferred to graphite conduits via thermal radiation. Inside the conduits, liquid tin is used as the heat transfer fluid. The tin is heated from 1900 °C to 2400 °C, transforming the energy input into sensible heat and increasing its enthalpy. The liquid tin is continuously pumped through the conduits and then conveyed to the graphite blocks in the storage unit. When the 2400 °C tin is pumped through the graphite blocks via conduits, it heats the graphite blocks from 1900 °C to 2400 °C via thermal radiation. Consequently, this cools the tin back to 1900 °C. The tin is then reheated by being pumped back through the resistance heaters. This process constitutes the charging process until the graphite blocks are heated back to peak temperature. The storage unit should have a sufficiently large thermal mass to enable storage unit to be charged for long periods with low heat loss.

The operating temperature and heat loss of the TEGS system is crucial design parameters, as lower temperatures result in lower capital cost per energy (CPE) due to reduced insulation requirements, while higher temperatures lead to higher capital costs per power (CPP) due to

[1] The PowerGenome Github library collects source data from EIA, NREL, and EPA and formats the input files for the CEM model. The GitHub library could with associated documentation could be found here: https://github.com/PowerGenome/PowerGenome

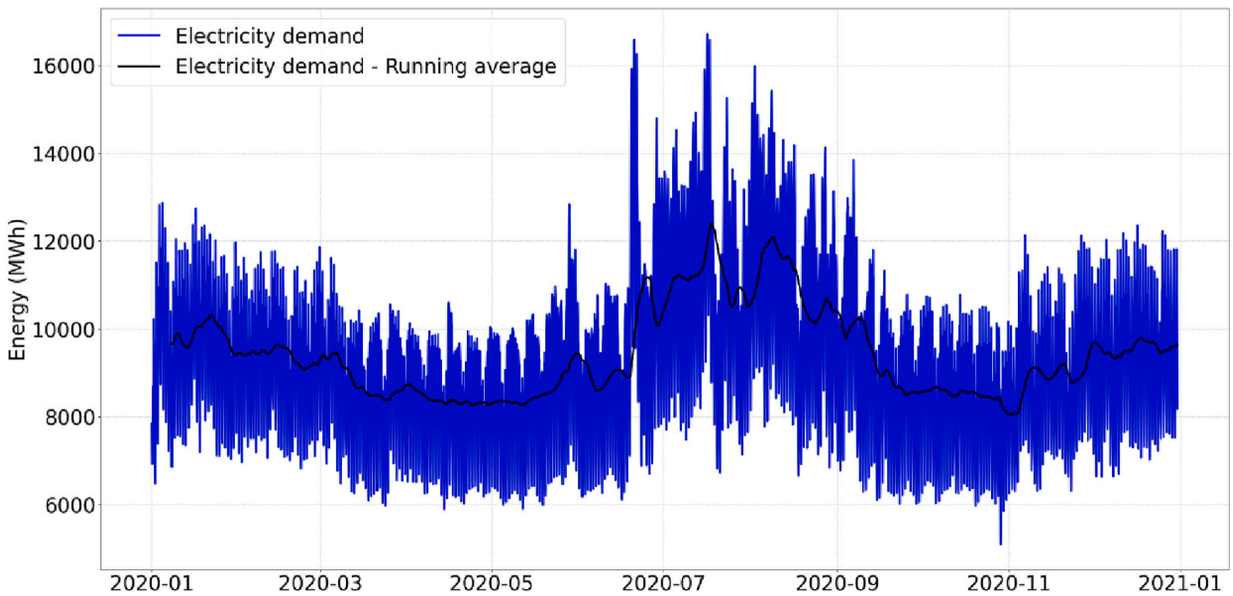


Fig. 1. Hourly electricity demand for the modeled power system.

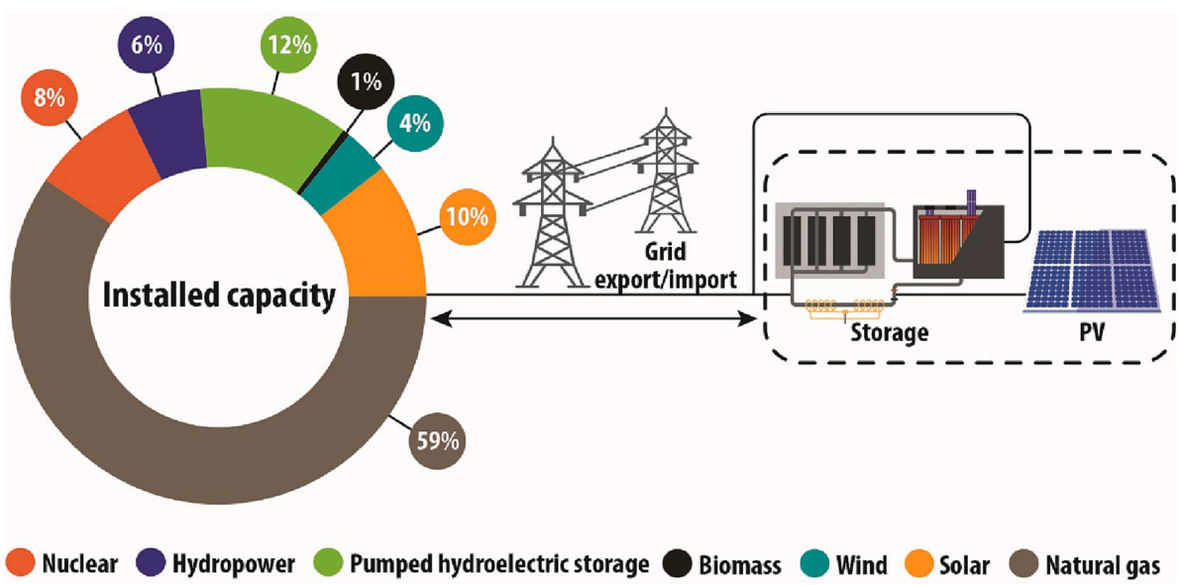


Fig. 2. The modeled New England grid zone is interconnected with a hypothetical PV + storage system. The existing power system is dominated by electricity generation from the NG. Solar and wind power represent a smaller share of electric power systems.

the need for more Thermophotovoltaic equipment. The Supplementary material section 1.3 presents an optimization procedure of the TEGS unit to identify the most cost-effective design when operating in the electric power system representing New England. It was found that optimizing the TEGS design with a daily heat loss of 1–3% and an operating temperature of 2400 °C proves to be the most cost-effective engineering design. Therefore, in this study, the modeled TEGS system has a daily heat loss of 1% and a temperature of 2400 °C.

During discharging, liquid tin is pumped through the graphite storage to a power block. The power block consists of graphite conduits with unit cells. Each unit cell of piping creates a rectangular cavity lined with tungsten foil. This is a diffusion barrier to prevent graphite deposition onto the TPV cells. Inside each cavity, the TPV cells can be lowered into the unit cell cavity. Here the TPV cells will be illuminated with the light emitted by the tungsten foil, which is heated by the light emitted by the graphite conduit carrying the tin. This net transfer of energy converts a large fraction (> 50%) of the energy to electricity, which causes the tin's temperature to decrease to 1900 °C before being pumped back to the graphite storage unit, where the tin is reheated again during the charging phase.

In this way, the TEGS is a rechargeable grid-scale thermal battery that can store energy as heat and supply electricity to the grid on demand, with an estimated RTE of 50%. The TPV conversion efficiency (i.e., the discharge efficiency from heat to electricity) entirely determines the RTE [17,18] . The charging efficiency (from heat to electricity) is assumed to be 100%.

2.2. Capacity expansion model (CEM) configuration and storage modeling

2.2.1. Capacity expansion model (CEM)

The analysis utilizes GenX [26], an electric power system CEM that evaluates a cost-optimal portfolio of electricity generation technologies, storage, and transmission to serve a given electricity demand. A detailed

description of all features the GenX model is described in Jenkins et al. [26]. The GenX CEM modeling procedure is subject to operational (electricity demand and generation) and policy (CO2 emission) constraints.

In this study, the operational constraints that are activated in the CEM are (1) the thermal generators' commitment on start-up/shut-down decisions, minimum and maximum up/down times (6 h), as well as hourly ramping limits for thermal generators (0.64, indicating the maximum increase and decrease in power output as a faction of the nameplate capacity), (2) transmission capacity limits between the existing electric grid and the hypothetical PV + TEGS system, (3) TEGS storage constraints on maximum hourly charge/discharge capacities and efficiency, stored energy, and daily heat loss, (4) Maximum capacity of the hypothetical PV plant that are coupled with the TEGS storage unit. The policy constraint activated in GenX in this study was the maximum limit on the allowed CO2 emissions in the grid. Here, the power system was modeled with a baseline scenario without any CO2 constraints (i.e., the model finds the cost-optimized electricity mix regardless of CO2 emissions) and with a scenario where the CO2 emissions are reduced by 50%. When the emissions are constrained by 50%, the electricity system must be less dependent on fossil-fuel-based technologies such as NG. That is, this requires to retire more of the current NG capacity in the grid and install more of the current VRE sources as well as utilize the hypothetical PV + TEGS system to a higher degree that has zero emissions associated with it.

To fully capture high-resolution temporal dependencies in the grid, the grid operation were modeled for each hour of the year. The resulting CEM configuration was solved as a mixed integer linear program (MILP). The Gurobi solver was used for the optimization problem in GenX as it provides the capability of solving MILP problems computationally efficiently [36]. The Gurobi optimization solver was applied using 16 cores with 128 GB RAM. All model scenarios were terminated with a 1% or lower optimality gap.

The objective function of the GenX model in this study is computed as the power system cost (PSC) grouped into the costs associated with the operation cost of the existing generators and cost associated with adding the hypothetical PV + TEGS system. The objective function in GenX is computed as

$$PSC = (Fix.Cost_{VRE} + Fix.Cost_{THERM} + Var.Cost + Start.Cost_{THERM}) + Inv.Cost_{TRANS} + Inv.Cost_{PV} + Inv.Cost_{TEGS} \quad (1)$$

here the $\mathrm{Fix.Cost_{VRE}}$ and $\mathrm{Fix.Cost_{THERM}}$ is the investment and Fixed O&M cost of the existing VRE and thermal capacity. The Var.Cost is the variable cost of generator dispatch, cost of non-served energy (periods where the electricity demand is not met), and cost of violating operating reserve requirements. The $\mathrm{Start.Cost_{THERM}}$ is the cost for startup and shutdowns for thermal power plants (NG). The $\mathrm{Inv.Cost_{TRANS}}$, $\mathrm{Inv.Cost_{PV}}$, and $\mathrm{Inv.Cost_{TEGS}}$ is the investment cost of adding the hypothetical system to the existing power grid.

2.2.2. Storage modeling

Three different TEGS sizes coupled with the 100 MW PV plant were modeled. The modeled energy storage capacities were as follows: 1) 400 MWh, 2) 600 MWh, and 3) 800 MWh. The sizes reflect the minimum TEGS storage capacity required to obtain the sufficient low capital cost of < U.S. $ 20/kWh for long-duration energy storage [17,18]. For each storage size, the charging capacity (i.e., the amount of energy that can be charged within one hour) is 100 MW, and the discharging capacity is in the range of [5, 100] MW. The storage unit is modeled to have a daily heat loss of 1%. As for other CEM studies evaluating storage [10], the storage capacity degradation or dynamic operation range (efficiency and capacity) is not modeled as it will significantly decrease the computational efficiency.

The storage configurations were modeled under the baseline case (with no CO2 reduction constraint) and 50% CO2 reduction scenarios. In total, 66 scenarios were modeled to address PV + TEGS energy availability with different storage sizes, discharge capacities, and CO2 constraints.

Since it is of interest to model the potential of utilizing TEGS to decarbonize future electric power systems, which are increasingly dependent on VRE technologies, the TEGS system is assumed to have a 50% RTE.

3. Experimental evaluations

In this study, it is of interest to assess the amount of time the hypothetical PV + TEGS system is available to the grid on demand. The Power availability factor (PAF) was computed to describe power availability. The PAF is computed as the percentage of time during the year the modeled PV + storage system can deliver at least a minimum quantity of power requested by the grid. Moreover, PAF allows the examination of the power availability of the combined power plant, as it measures how often it can supply a minimum amount of power to the grid. A power plant with a 100% PAF can always provide a given minimum amount of power to the grid.

In the case of solar PV, the electricity generation suddenly drops when the sun no longer shines (owing to cloud cover or when the sun sets). This sudden drop in electricity generation reduces the PAF because the PV plant no longer generates at the rated power. Here, storage units can be used to charge whenever there is a low net demand for power in the grid and to discharge when there is a higher demand for power. Fig. 3 illustrates an example of how the storage unit can be used to shape the output to provide constant baseload power. Once the PV plant starts generating electricity over derated power (e.g., 20 MW), excess electricity is used to charge the storage unit. When the solar plant generates less electricity than the derated power, the storage unit starts discharging to satisfy the demand for electricity.

Fig. 3 illustrates how the combined PV and storage system can provide constant baseload power, disregarding the electricity demand in the grid. However, when the system is connected to an electricity grid, it becomes significantly more complex. The system should not provide constant baseload power to the grid but should be able to supply the requested power to the grid system operator.

In this study, the periods where the hypothetical PV + storage system cannot provide the requested electricity to the grid on demand are detected.

The grid is requesting electricity from the hypothetical PV + storage system in all periods when there is no excess electricity in the grid that can be used to charge the TEGS unit (i.e., no export from the grid to charge the TEGS unit). The unwanted periods when the electricity grid request (i.e., no excess electricity that can be exported from the grid to the PV + storage system) power arise for the following reasons: 1) The PV system does not deliver the required power, and 2) The TEGS system cannot discharge the requested power as the State of Charge (SOC) is already zero. Such critical periods can be calculated as:

$$\text{Percentage Not Available} = \frac{\text{When}\big(\big(\mathrm{PV_{gen}} + \mathrm{TEGS_{discharge}} < \text{Derated power}\big) + (\text{SOC} = 0)\big)}{8760 \text{ hours}} \times 100 \quad (1)$$

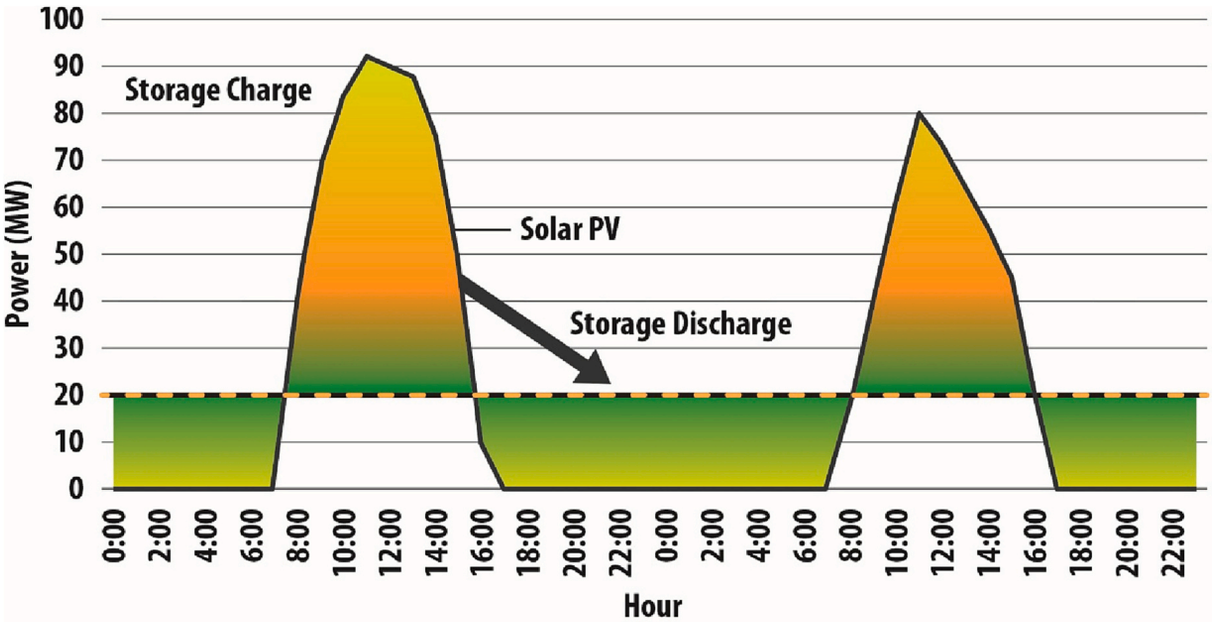


Fig. 3. Example illustration where storage is used to provide a constant baseload power to the grid by charging when there is excess electricity and discharging when the PV plant does not generate electricity.

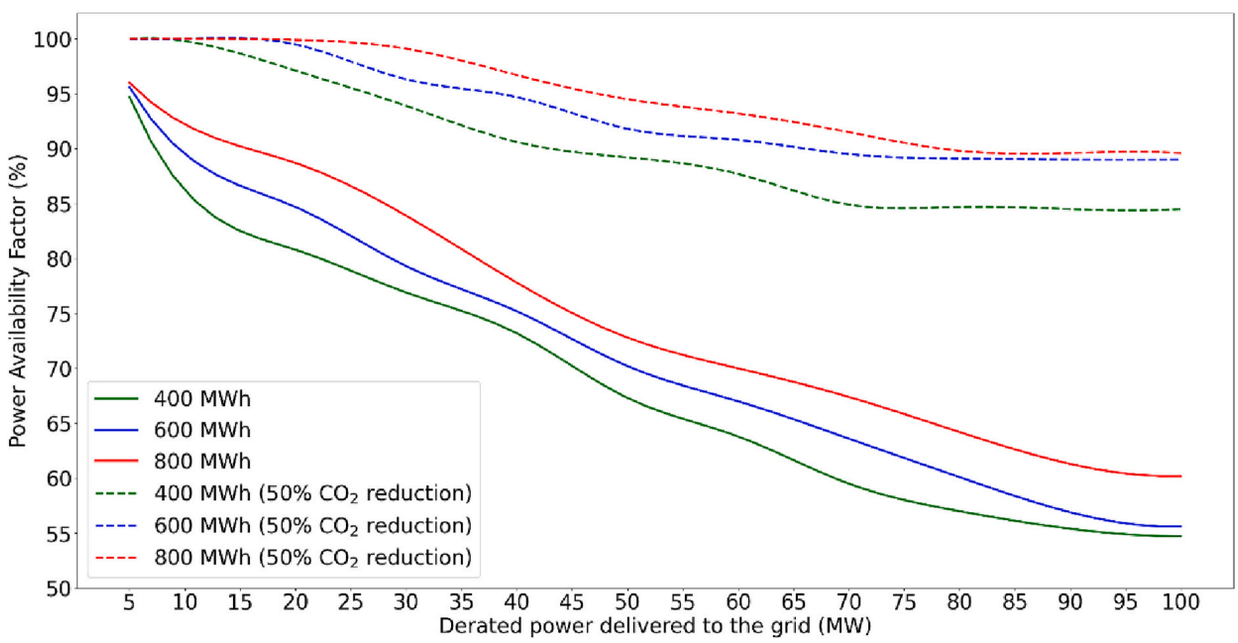


Fig. 4. Percentage of time during the year when the hypothetical system can deliver the requested power to the grid.

The system Percentage Not Available (PNA) gives information about the percentage of time during a full year the system cannot provide the requested electricity to the grid. On contrary, the PAF will be calculated as:

$$\mathrm{PAF} = 1 - \mathrm{PNA} \tag{2}$$

4. Results and discussions

The resulting yearly PAF of the hypothetical PV + TEGS system as a function of different discharge powers is provided in Fig. 4. Here, the yearly PAF is computed from Eq. (1) and Eq. (2) using the resulting data from the GenX optimization schedule of the New England electric grid region. The more the system is derated, the more often it can deliver the required power. This is reasonable because the more the system is derated, the more the storage system can discharge at a lower-rated output for a longer period. The PV + TEGS system can contribute 5 MW to the grid for about 95% of the hours throughout the year if the storage system has a discharge capacity of 5 MW.

For the 600 MWh storage unit, the PV + TEGS system can contribute 5 MW to the grid for about 95% of the hours throughout the year if the storage system has a discharge capacity of 5 MW. The results for the other storage sizes are similar, with a PAF of 94% and 96% for the 400 MWh and 800 MWh units, respectively. If the discharge capacity is 100 MW, the PV + TEGS system can deliver 100 MW to the grid approximately 55–60% of the hours during the year for all storage sizes (400 MWh, 600 MWh, and 800 MWh).

Interestingly, there is a large difference between the scenarios with and without CO_2 reduction constraints. The electric power system dynamics change completely once CO_2 emissions were reduced by 50%. In this case, the increased retirement of the NG makes the grid more dependent on the hypothetical PV + TEGS system, which results in the system supplying the necessary power 100% of the time for a derated power between 5 MW and 20 MW for the TEGS unit with a storage capacity of 600 MWh and 800 MWh. This is remarkably higher than the baseline case, where the system cannot deliver the required power 5–15% of the time for such derated powers. When modeling the PAF for the PV + TEGS system at the GW scale (see the Supplementary material), similar results are obtained. The PAF increases when lowering the derated power in both the baseline and the CO_2 reduction scenarios. In addition, the PAF is significantly higher when modeling the CO_2 reduction scenario compared to the baseline scenario.

The large (30% higher PAF on average in the 50% CO2 reduction scenario) difference between the CO2 reduction scenarios indicates that the electric power system dynamic changes significantly if the grid must be less dependent on NG.

Now the hypothetical PV + storage system plays a more important role in the grid because it does not emit any CO2, and as such, the cost-minimization schedule of the CEM optimizes the grid to ensure that the PV + storage system can deliver the requested power to the grid more often during the year.

Fig. 5 selects one derate level from Fig. 4 to investigate how the PAF changes with different CO2 reduction scenarios. More specifically, Fig. 5 illustrates how the PAF changes with decreased CO2 emissions for a storage unit of 600 MWh that discharges 20 MW to the grid. Clearly, reducing CO2 emissions results in a higher PAF. The maximum PAF was achieved at 50% CO2 reduction. Reducing CO2 emissions requires the power system to be less dependent on fossil fuel technologies, such as NG, and thus must rely more on the power served by the PV + storage system.

Modeling scenarios with >50% CO2 reduction results in an infeasible solution with the CEM optimization. That is, the objective function of the GenX model to cost-optimize the portfolio of electric power generation to serve the demand for electricity is not fullfilled. Therefore, to enable the possibility of further reducing the CO2 emissions in the modeled grid, the existing portfolio of electric generation technologies must be expanded and include more emission-free technologies that can serve the demand for electricity.

Fig. 6 shows how the hypothetical system operates in the different emissions scenarios (baseline and 50% CO2 reduction). Here, an example of the 600 MWh storage unit with a 20 MW discharge capacity that has a yearly PAF of 79% (baseline) and 100 (50% CO2 reduction) from Fig. 4 is illustrated. The weekly examples show the hourly operation of the hypothetical system during a typical winter week in January. The uppermost and lowermost Figure show how the system operates in scenarios with and without constraining CO2 emissions. From both graphs, TEGS is used frequently to discharge power to the grid whenever there is low electricity generation from the solar PV, which reduces the intermittency problem of PVs by increasing the number of hours the hypothetical system can deliver the required power to the grid. In addition, TEGS also charges power from the grid whenever there is a drop in the demand to increase the SOC, which illustrates the benefit of using storage units that are coupled with VRE resources.

The yellow area highlights a critical period in the uppermost graph. The critical period is defined as the incidents where the grid has an increasing net load (blue line), but the system cannot deliver the requested power because there is no generation from the PV system, and the storage unit cannot discharge the required power to the grid because the SOC is already zero. The net load is given as the total electricity demand subtracted by the electricity generation from solar and wind power.

However, considering the 50% CO_2 reduction scenario, the system interacts differently with the grid, and it is clear that the CEM optimizes the hypothetical system to have more energy available at more times because the grid now is more dependent on the hypothetical system (because the grid can use less NG). For the particular example week, the PV + storage system can deliver the requested power to the grid at all times for the 50% CO_2 reduction scenario.

Like the winter week example shown in Figs. 6, 7 shows how the hypothetical system operates with the grid during a typical summer week. Here, thanks to the higher solar availability, the system can deliver most of the required power to the grid using only the PV plant, and the storage system is used less frequently. However, for the baseline case, there are several periods in which the grid requests energy and the storage unit cannot provide sufficient power to the grid because the SOC is already zero. This is because the grid is mainly dependent on NG, which can supply power whenever there is low solar availability, and the TEGS system is only used to provide additional peak power when the demand suddenly increases.

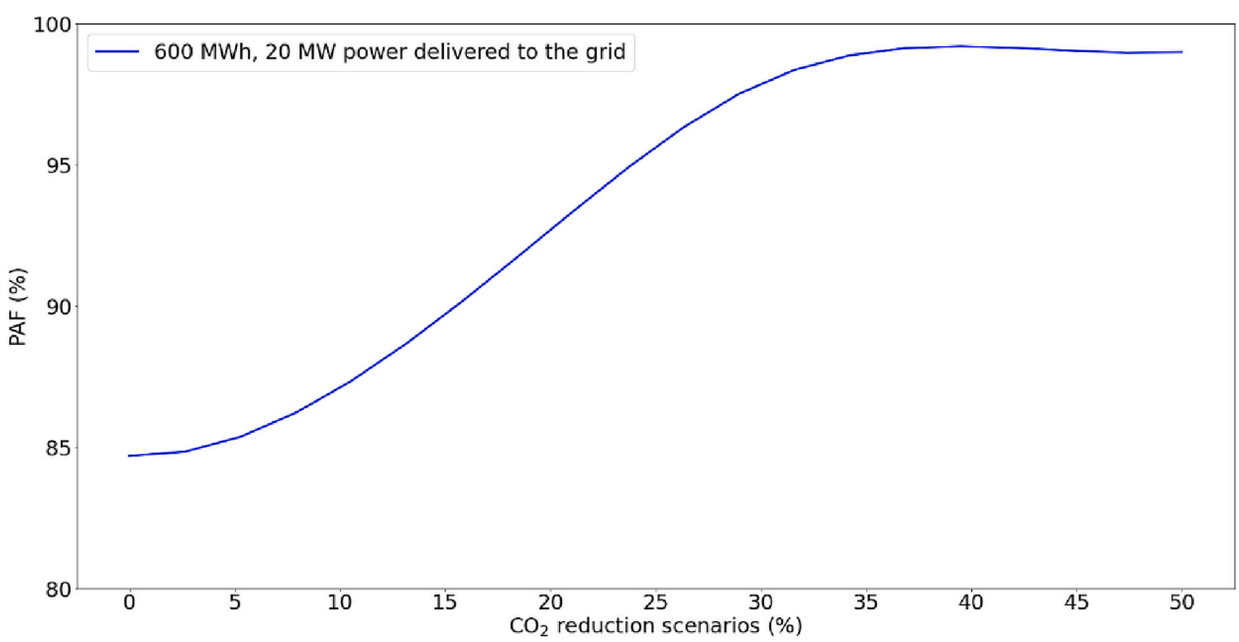


Fig. 5. Percentage of time the hypothetical system can deliver the requested 20 MW power to the grid as a function of CO_2 reduction scenarios.

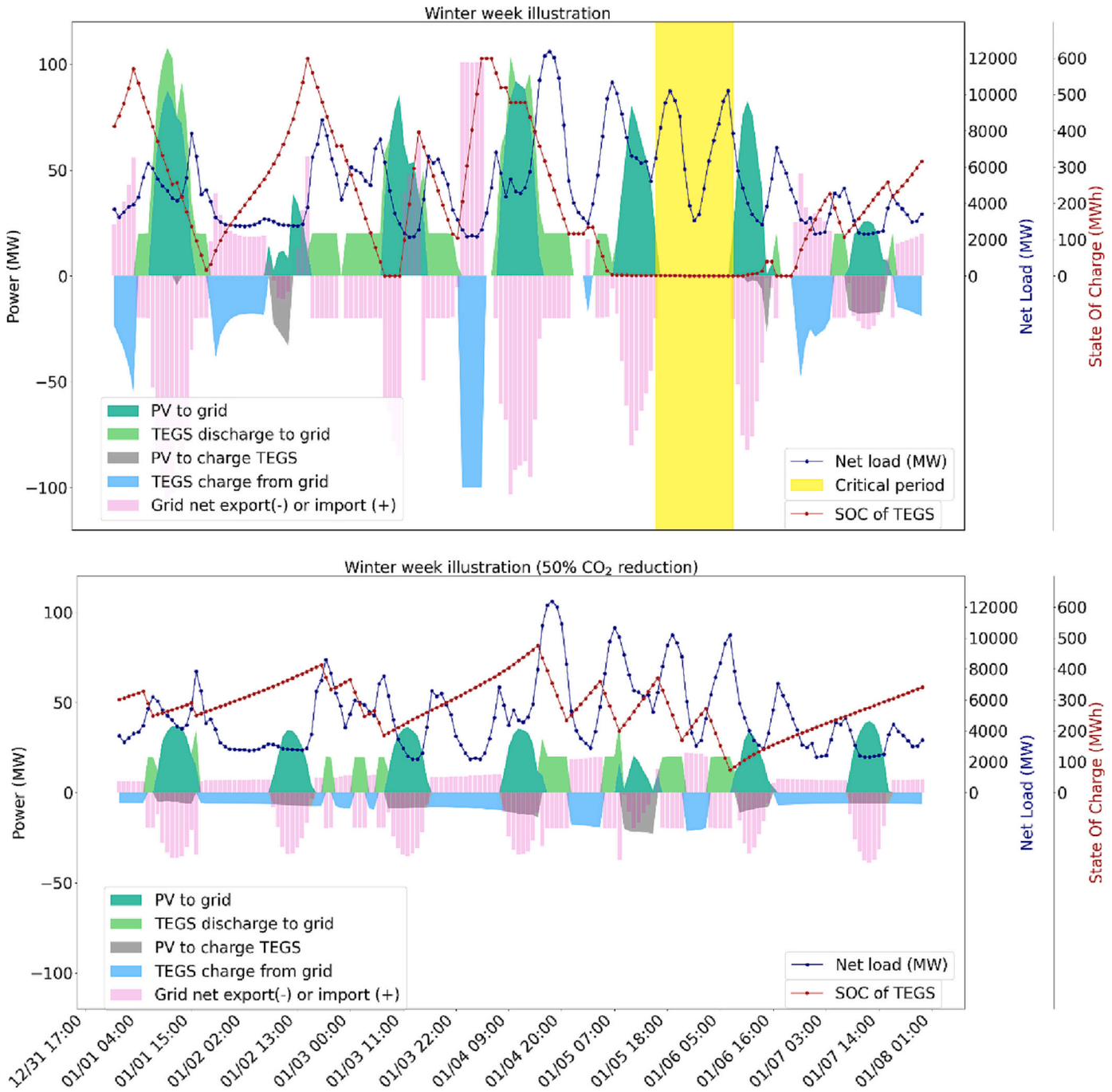


Fig. 6. Winter week illustrations for a 100 MW PV plus a 600 MWh TEGS system that can discharge 20 MW. Baseline case without constraining CO_2 emissions (uppermost graph) and future scenario with reduced emissions lowermost.

Under the 50% CO_2 reduction scenario, the grid is much more dependent on the power from the storage unit whenever there is no PV generation. It is clear that instead of providing peak power to the grid, the PV system is used to charge the TEGS unit to a higher degree to ensure that the SOC is never zero and thus can discharge the derated power to the grid at all times when the PV system does not generate electricity.

5. Conclusion

In this study, the potential of using energy storage to tackle the intermittency problem of VRE sources by increasing the dispatchability of a hypothetical PV plant were analyzed. An existing electricity grid region in North America were modeled using a CEM and investigated how different storage configurations can reduce the number of periods in which a hypothetical PV + storage system cannot provide the required power to the grid.

Because of the high capital cost of electrochemical batteries, a TES technology with a projected capital cost that fulfills the requirements ($<$ US$ 20 kWh^{-1}) to enable full decarbonization of the grid was considered. The energy availability of the hypothetical system was modeled under different storage sizes and discharge capacities. Additionally, the optimization schedule was repeated under a hypothetical future scenario in which CO_2 emissions were constrained to be reduced by 50%. In total, 66 different scenarios were modeled. To capture the high-resolution dependencies in the electricity generation balance, a full year with hourly resolution was optimized using the CEM.

The results support the added value of using storage to increase the

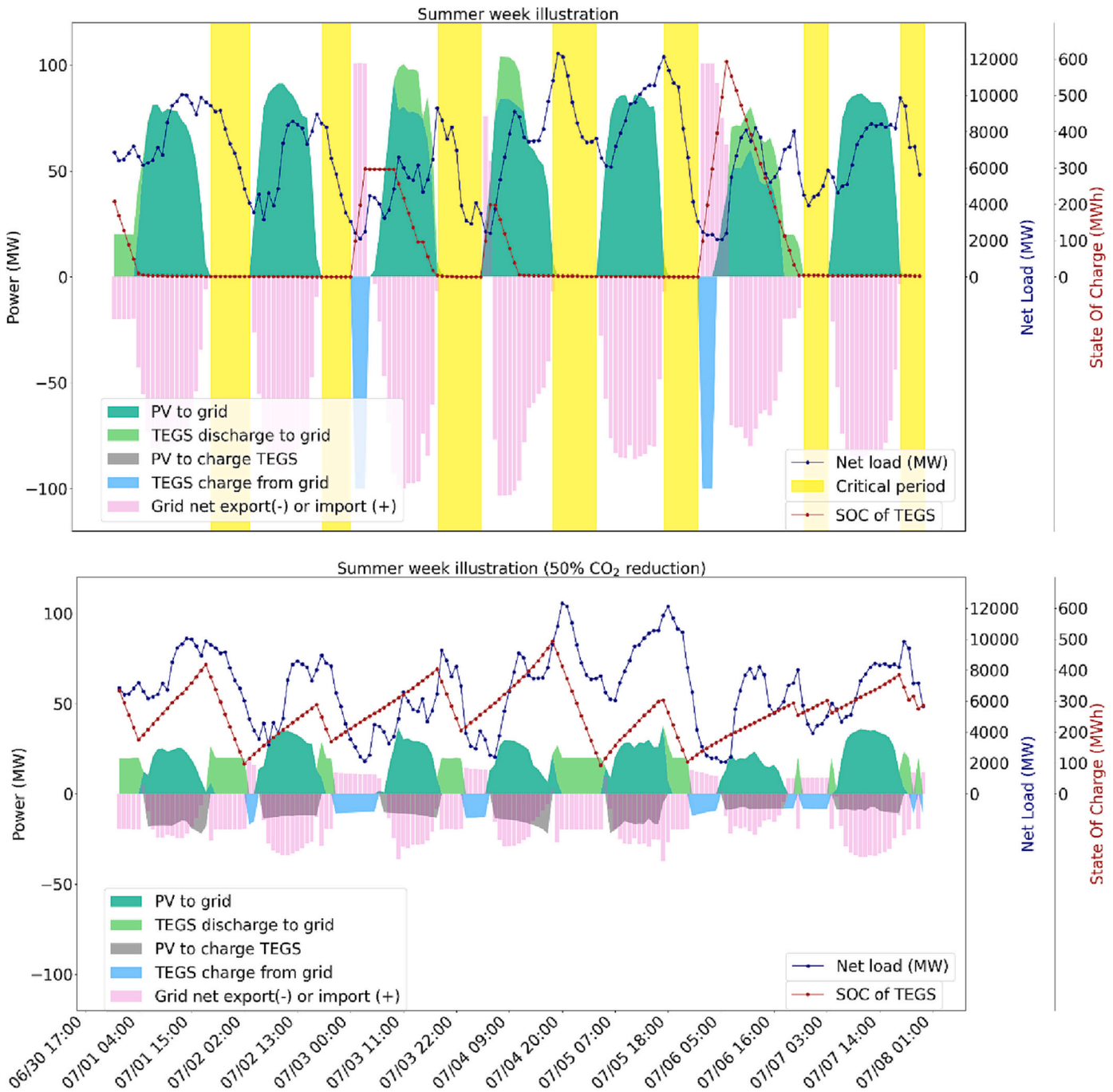


Fig. 7. Illustrations of a typical summer week for a 100 MV PV plus 600 MWh TEGS system that can discharge 20 MW. Baseline case without constraining the CO2 emissions (uppermost graph), and future scenario with reduced emissions lowermost.

dispatchability of PV, as it significantly increases the PAF compared to PV systems alone. The percentage of time during the year the system could deliver the required power to the grid increased when the discharge capacity of the system decreased. In addition, increasing the storage size increases the energy availability, as more energy can be stored and thereafter discharged over a longer period when there is a demand for electricity in the grid. The findings were consistent when the PV plus TEGS system were evaluated at both the megawatt and gigawatt scale.

Interestingly, there was a significant change in the electricity grid generation dynamics when the CO_2 emissions were reduced by 50%. Here, because the grid can no longer rely on the same share of NG technology, the most cost-efficient grid is achieved when the PV + TEGS system is utilized to a higher degree, as these technologies do not emit any CO_2.

This shows that decreasing the maximum allowed GHG emissions in the grid significantly increases the value of using storage to increase the dispatchability of PV systems.

The study findings could provide increased knowledge to power system planners regarding how adding PV + storage systems to existing grids can contribute to the efficient stepwise decarbonization of power systems.

5.1. Limitations and suggested future research

This study presented an idealized representation of an existing grid (i.e., "brownfield" CEM approach) in the New England grid region. However, the authors are fully aware that the grid representation might

not fully capture all details of the existing grid, and there can be differences (sizes of the power plants, electricity demand on the grid, share of the existing generation technologies) between our abstract grid representation and the current real-world grid that ISO New England operates. In addition to the transmission line between the existing grid and hypothetical PV + storage system, the current grid were modeled as a single-zone grid region without considering transmission losses or congestion between generators and demand.

This study's major objectives are to evaluate a hypothetical PV + storage system's power availability and discuss the significance of integrating such technologies to assure a successful step-by-step decarbonization of the electric grid. Therefore, modeling the transmission lines between the existing generators is not considered, as it is outside the scope of this study and will significantly increase the computational intensity of the CEM.

The CEM is fully deterministic, assumes perfect foresight in planning and operational decisions, and does not account for uncertainty in VRE generation [39–42]. Therefore, this study does not aim to be used as a power planning tool for ISOs to assess the PAF of PV + storage systems in the day-ahead electricity market, but shows how storage, in general, will be a valuable technology to address the intermittency issue of VRE. A suggested future study is to frame the CEM to account for the uncertainty regarding the expected electricity generation from VRE sources. This will allow the use of the CEM as a decision-making tool for optimizing the management of the electricity grid in the day-ahead market.

Credit author statement

Odin Foldvik Eikeland performed all analyzes and had the main responsibility for writing the manuscript. Colin C. Kelsall and Kyle Buznitsky contributed to the inception of the study and discussions during the analysis. Shomik Verma provided feedback on the draft of the paper, contributed to the discussions of the results, and provided supervision on how to do the modeling on the computing cluster. Filippo Maria Bianchi contributed by providing feedback on the draft and supervising how to structure the manuscript. Matteo Chiesa assisted in the analysis and interpretation of the results and contributed to the inception of the manuscript. Asegun Henry supervised the inception of the study and provided feedback and guidance in the analysis and interpretation of the results.

Odin Foldvik Eikeland wrote the manuscript with input from all authors.

CRediT authorship contribution statement

Odin Foldvik Eikeland: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Colin C. Kelsall:** Methodology, Investigation, Conceptualization. **Kyle Buznitsky:** Data curation, Formal analysis, Investigation. **Shomik Verma:** Investigation, Formal analysis, Conceptualization. **Filippo Maria Bianchi:** Writing – original draft, Supervision, Investigation. **Matteo Chiesa:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Asegun Henry:** Writing – review & editing, Writing – original draft, Supervision, Software, Resources, Project administration, Methodology, Investigation, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

O.F.E. and M.C. acknowledge the support from the research project "Transformation to a Renewable & Smart Rural Power System Community (RENEW)", connected to the Arctic Centre for Sustainable Energy (ARC) at UiT-the Arctic University of Norway through Grant No. 310026. We thank Maritsa Kissamitaki for designing Fig. 2 and Fig. 3.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.apenergy.2023.121572.

References

[1] Remi Eriksen DE, Haugen Ulrike, Hodne Trond, Hovem Liv, Huang Jin James, Bakken Bent Erik, et al. Energy Transition Outlook 2022, A global and regional forecast to 2050. Oslo, Norway: DNV AS; 2022.
[2] Gielen D, Boshell F, Saygin D, Bazilian MD, Wagner N, Gorini R. The role of renewable energy in the global energy transformation. Energ Strat Rev 2019;24:38–50. https://doi.org/10.1016/j.esr.2019.01.006.
[3] Apostoleris H, Sgouridis S, Stefancich M, Chiesa M. Utility solar prices will continue to drop all over the world even without subsidies. Nat Energy 2019;4(10):833–4. https://doi.org/10.1038/s41560-019-0481-4.
[4] Braff WA, Mueller JM, Trancik JE. Value of storage technologies for wind and solar energy. Nat Clim Change 2016;6(10):964–9.
[5] Ziegler MS, et al. Storage requirements and costs of shaping renewable energy toward grid decarbonization. Joule 2019;3(9):2134–53.
[6] Sepulveda NA, Jenkins JD, Edington A, Mallapragada DS, Lester RK. The design space for long-duration energy storage in decarbonized power systems. Nat Energy 2021;6(5):506–16. https://doi.org/10.1038/s41560-021-00796-8.
[7] Denholm P, Mai T. Timescales of energy storage needed for reducing renewable energy curtailment. Renew Energy 2019;130:388–99. https://doi.org/10.1016/j.renene.2018.06.079.
[8] Shaner MR, Davis SJ, Lewis NS, Caldeira K. Geophysical constraints on the reliability of solar and wind power in the United States. Energ Environ Sci 2018;11(4):914–25.
[9] Gandhi K, Apostoleris H, Sgouridis S. Catching the hydrogen train: economics-driven green hydrogen adoption potential in the United Arab Emirates. Int J Hydrogen Energy 2022;47(53):22285–301. https://doi.org/10.1016/j.ijhydene.2022.05.055.
[10] Mallapragada DS, Sepulveda NA, Jenkins JD. Long-run system value of battery energy storage in future grids with increasing wind and solar generation. Appl Energy 2020;275:115390. https://doi.org/10.1016/j.apenergy.2020.115390.
[11] Xu X, Vignarooban K, Xu B, Hsu K, Kannan AM. Prospects and problems of concentrating solar power technologies for power generation in the desert regions. Renew Sustain Energy Rev 2016;53:1106–31. https://doi.org/10.1016/j.rser.2015.09.015.
[12] Kennedy KM, et al. The role of concentrated solar power with thermal energy storage in least-cost highly reliable electricity systems fully powered by variable renewable energy. Adv Appl Energy 2022;6:100091. https://doi.org/10.1016/j.adapen.2022.100091.
[13] Albertus P, Manser JS, Litzelman S. Long-duration electricity storage applications, economics, and technologies. Joule 2020;4(1):21–32. https://doi.org/10.1016/j.joule.2019.11.009.
[14] P. Eds Core Writing Team R. K, Meyer LA. IPCC Climate Change 2014: Synthesis report. IPCC; 2014.
[15] U. E. P. A. (EPA). Inventory of US greenhouse gas emissions and sinks: 1990–2009. Washington. 2011.
[16] Henry A, Prasher R, Majumdar A. Five thermal energy grand challenges for decarbonization. Nat Energy 2020;5(9):635–7. https://doi.org/10.1038/s41560-020-0675-9.
[17] Amy C, Seyf HR, Steiner MA, Friedman DJ, Henry A. Thermal energy grid storage using multi-junction photovoltaics. Energ Environ Sci 2019;12(1):334–43. https://doi.org/10.1039/C8EE02341G.
[18] Kelsall CC, Buznitsky K, Henry A. Technoeconomic analysis of thermal energy grid storage using graphite and tin. arXiv 2021. https://doi.org/10.48550/arXiv.2106.07624.
[19] LaPotin A, et al. Thermophotovoltaic efficiency of 40%. Nature 2022;604(7905):287–91. https://doi.org/10.1038/s41586-022-04473-y.
[20] Schill W-P, Zerrahn A. Long-run power storage requirements for high shares of renewables: results and sensitivities. Renew Sustain Energy Rev 2018;83:156–71. https://doi.org/10.1016/j.rser.2017.05.205.
[21] de Sisternes FJ, Jenkins JD, Botterud A. The value of energy storage in decarbonizing the electricity sector. Appl Energy 2016;175:368–79. https://doi.org/10.1016/j.apenergy.2016.05.014.

[22] Heuberger CF, Staffell I, Shah N, Dowell NM. A systems approach to quantifying the value of power generation and energy storage technologies in future electricity networks. Comp Chem Eng 2017;107:247–56. https://doi.org/10.1016/j.compchemeng.2017.05.012.
[23] Sepulveda NA, Jenkins JD, de Sisternes FJ, Lester RK. The role of firm low-carbon electricity resources in deep decarbonization of power generation. Joule 2018;2(11):2403–20. https://doi.org/10.1016/j.joule.2018.08.006.
[24] Frew BA, Becker S, Dvorak MJ, Andresen GB, Jacobson MZ. Flexibility mechanisms and pathways to a highly renewable US electricity future. Energy 2016;101:65–78. https://doi.org/10.1016/j.energy.2016.01.079.
[25] Liu H, Brown T, Andresen GB, Schlachtberger DP, Greiner M. The role of hydro power, storage and transmission in the decarbonization of the Chinese power system. Appl Energy 2019;239:1308–21. https://doi.org/10.1016/j.apenergy.2019.02.009.
[26] Jafari M, Korpås M, Botterud A. Power system decarbonization: impacts of energy storage duration and interannual renewables variability. Renew Energy 2020;156:1171–85. https://doi.org/10.1016/j.renene.2020.04.144.
[27] Schleifer AH, Murphy CA, Cole WJ, Denholm P. Exploring the design space of PV-plus-battery system configurations under evolving grid conditions. Appl Energy 2022;308:118339.
[28] Limpens G, Moret S, Jeanmart H, Maréchal F. EnergyScope TD: a novel open-source model for regional energy systems. Appl Energy 2019;255:113729. https://doi.org/10.1016/j.apenergy.2019.113729.
[29] Jenkins JDS. Enhanced decision support for a changing electricity landscape: the GenX configurable electricity resource capacity expansion model [Online]. Available: https://energy.mit.edu/wp-content/uploads/2017/10/Enhanced-Decision-Support-for-a-Changing-Electricity-Landscape.pdf; 2017.
[30] Guerra OJ, Zhang J, Eichman J, Denholm P, Kurtz J, Hodge B-M. The value of seasonal energy storage technologies for the integration of wind and solar power. Energ Environ Sci 2020;13(7):1909–22.
[31] Schulthoff M, Rudnick I, Bose A, Gençer E. Role of hydrogen in a low-carbon electric power system: a case study. Front Energy Res 2020. https://doi.org/10.3389/fenrg.2020.585461.
[32] Bompard E, et al. An electricity triangle for energy transition: application to Italy. Appl Energy 2020;277:115525. https://doi.org/10.1016/j.apenergy.2020.115525.
[33] Cole WJ, et al. Quantifying the challenge of reaching a 100% renewable energy power system for the United States. Joule 2021;5(7):1732–48. https://doi.org/10.1016/j.joule.2021.05.011.
[34] Denholm P, et al. The challenges of achieving a 100% renewable electricity system in the United States. Joule 2021;5(6):1331–52. https://doi.org/10.1016/j.joule.2021.03.028.
[35] Baik E, et al. What is different about different net-zero carbon electricity systems? Energy Clim Change 2021;2:100046. https://doi.org/10.1016/j.egycc.2021.100046.
[36] Creutzig F, Agoston P, Goldschmidt JC, Luderer G, Nemet G, Pietzcker RC. The underestimated potential of solar energy to mitigate climate change. Nat Energy 2017;2(9):17140. https://doi.org/10.1038/nenergy.2017.140.
[37] Tyra BC. Christopher; Harrison, Eric; Gorski, Alexander, "electric power monthly, march 2023,". U.S. Energy Information Administration; 2023.
[39] Zhang Y, Wang J, Wang X. Review on probabilistic forecasting of wind power generation. Renew Sustain Energy Rev 2014;32:255–70. https://doi.org/10.1016/j.rser.2014.01.033.
[40] Salinas D, Flunkert V, Gasthaus J, Januschowski T. DeepAR: probabilistic forecasting with autoregressive recurrent networks. Int J Forecast 2020;36(3):1181–91. https://doi.org/10.1016/j.ijforecast.2019.07.001.
[41] Mashlakov A, Kuronen T, Lensu L, Kaarna A, Honkapuro S. Assessing the performance of deep learning models for multivariate probabilistic energy forecasting. Appl Energy 2021;285:116405. https://doi.org/10.1016/j.apenergy.2020.116405.
[42] Eikeland OF, Hovem FD, Olsen TE, Chiesa M, Bianchi FM. Probabilistic forecasts of wind power generation in regions with complex topography using deep learning methods: an arctic case. Energ Conver Manage X 2022;15:100239. https://doi.org/10.1016/j.ecmx.2022.100239.

Power Availability of PV plus Thermal Storage units in real-world electric power grids

Odin Foldvik Eikeland[1,2], Colin C. Kelsall[1], Kyle Buznitsky[1], Shomik Verma[1], Filippo Maria Bianchi[3], Matteo Chiesa[1,2,4,*], & Asegun Henry[1,*]

[1] Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA.
[2] Department of Physics and Technology, UiT – the Arctic University of Norway, 9037 Tromsø, Norway.
[3] Department of Mathematics and Statistics, UiT – the Arctic University of Norway Technology, 9037 Tromsø, Norway.
[4] Laboratory for Energy and NanoScience (LENS), Khalifa University of Science and Technology, Masdar Institute Campus, 127788 Abu Dhabi, United Arab Emirates
[*] Corresponding authors: mchiesa@mit.edu, ase@mit.edu

1. Supplementary material

1.1. Extended results and discussions

1.1.1 Hourly grid operations during spring and autumn weeks

To access the performance of the PV + TEGS system in different seasons, two weekly examples of the hourly operation for the spring and the autumn season are included in the Supplementary material. The weekly examples illustrate the hourly operation of the PV + TEGS system where the PV system has a capacity of 100 MW, and the TEGS unit has a 600 MWh storage capacity with a 20 MW discharge capacity. Like in the main manuscript, the uppermost and lowermost figure show how the system operates in scenarios with and without constraining CO_2 emissions.

Figure 1 shows how the hypothetical system operates during a spring week. Similar to Figure 5 and Figure 6 in the main manuscript, the weekly examples show that the hypothetical system operates differently when the CO_2 emissions are constrained to be reduced by 50%. In the baseline scenario with no CO_2 emission constraints, there is a critical period where the grid has increased electricity demand (blue line), but the PV + TEGS system cannot deliver the requested power to the grid. Considering the 50% CO_2 reduction scenario, there are no critical periods, and the hypothetical PV + TEGS system can always provide the requested power to the grid.

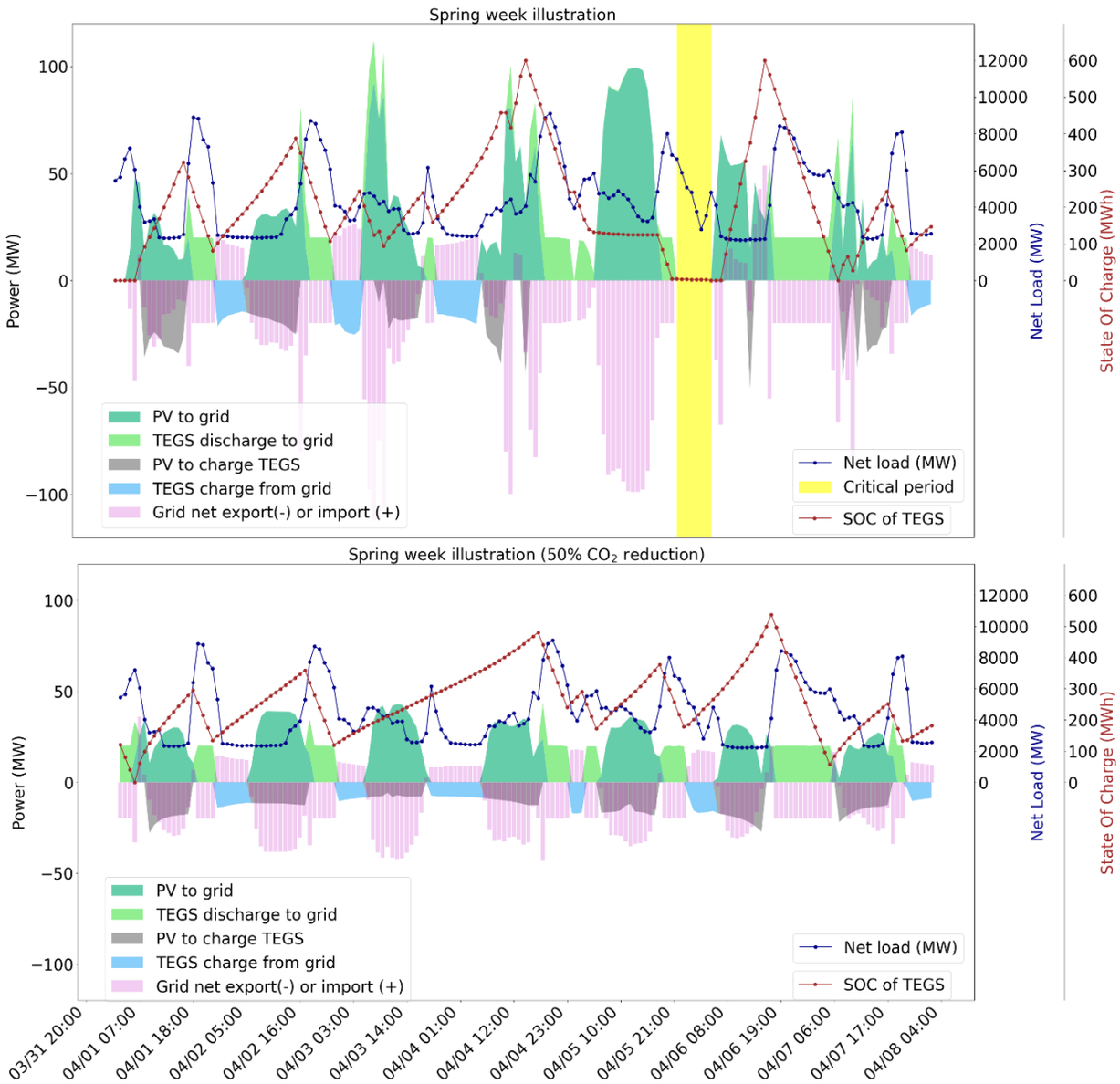


Figure 1: Spring week illustrations for a 100 MW PV plus a 600 MWh TEGS system that can discharge 20 MW. Baseline case without constraining CO_2 emissions (uppermost graph) and a scenario with reduced emissions lowermost.

Similar to Figure 1, Figure 2 shows how the hypothetical system operates during an autumn week. Clearly, there are several periods during the autumn for the baseline scenario where the PV + TEGS system cannot serve the grid with the requested power since the electricity generation from the PV system is zero, and the State of Charge (SOC) of the TEGS unit is zero. However, when the CO_2 emissions are constrained by 50%, the CEM optimizes the portfolio of the generation technologies to ensure that the grid can always get the requested power from the hypothetical PV + TEGS system.

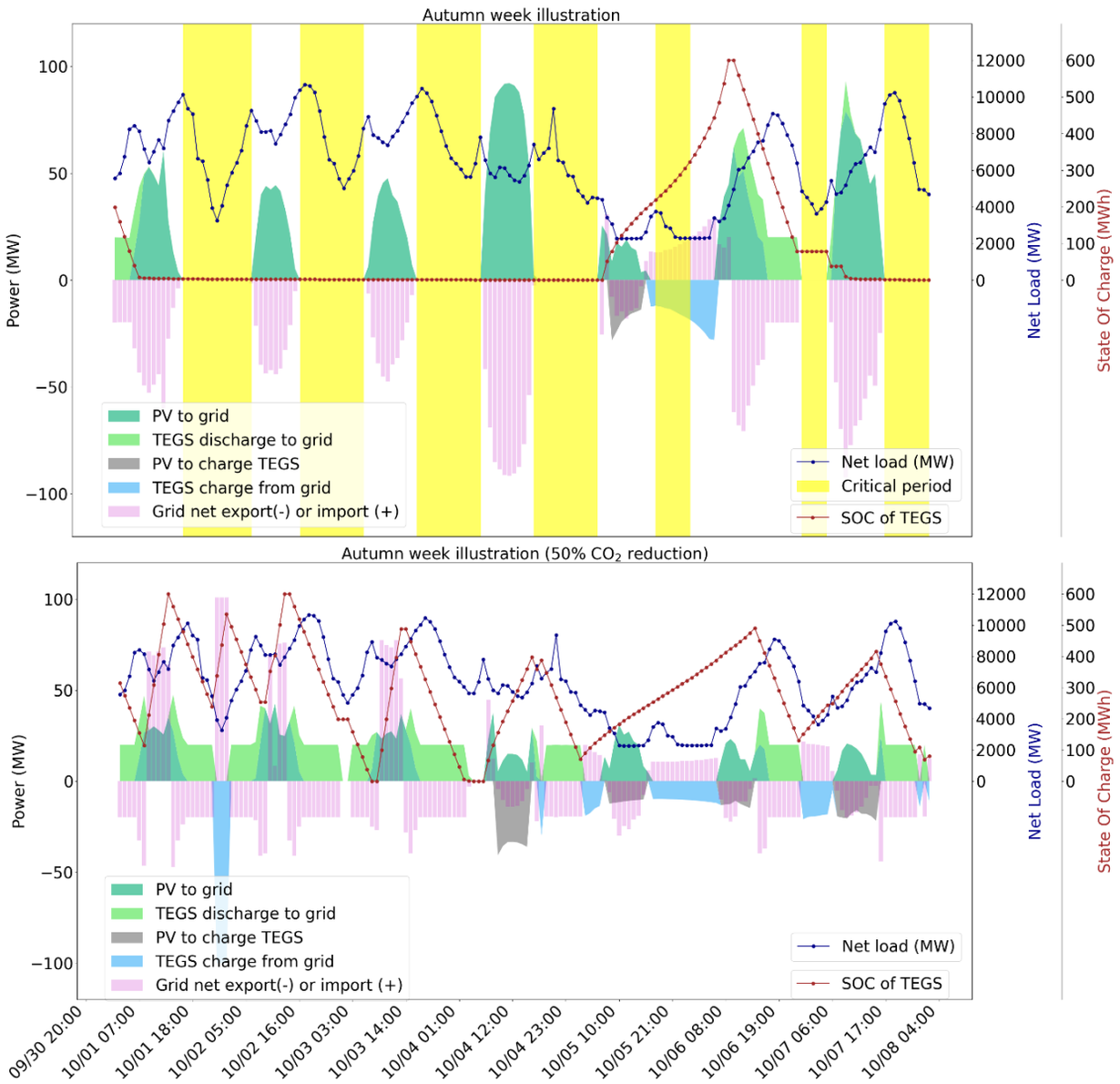


Figure 2: Autumn week illustrations for a 100 MW PV plus a 600 MWh TEGS system that can discharge 20 MW. Baseline case without constraining CO_2 emissions (uppermost graph) and a scenario with reduced emissions lowermost.

1.1.2 Power availability under low electricity generation from solar PV

To access the performance of the PV + TEGS system under a week with low electricity generation from solar energy, Figure 3 illustrates another example week from the winter season is included. Interestingly, the PV + TEGS system can serve the requested power from the grid almost at all times in the baseline scenario, except for one critical period. This is due to lower electricity demand in the grid, which requires less power supply from the hypothetical PV + TEGS system to meet the requested demand for electricity. For the 50% CO_2 reduction scenario, the PV + TEGS system can always meet the requested electricity demand.

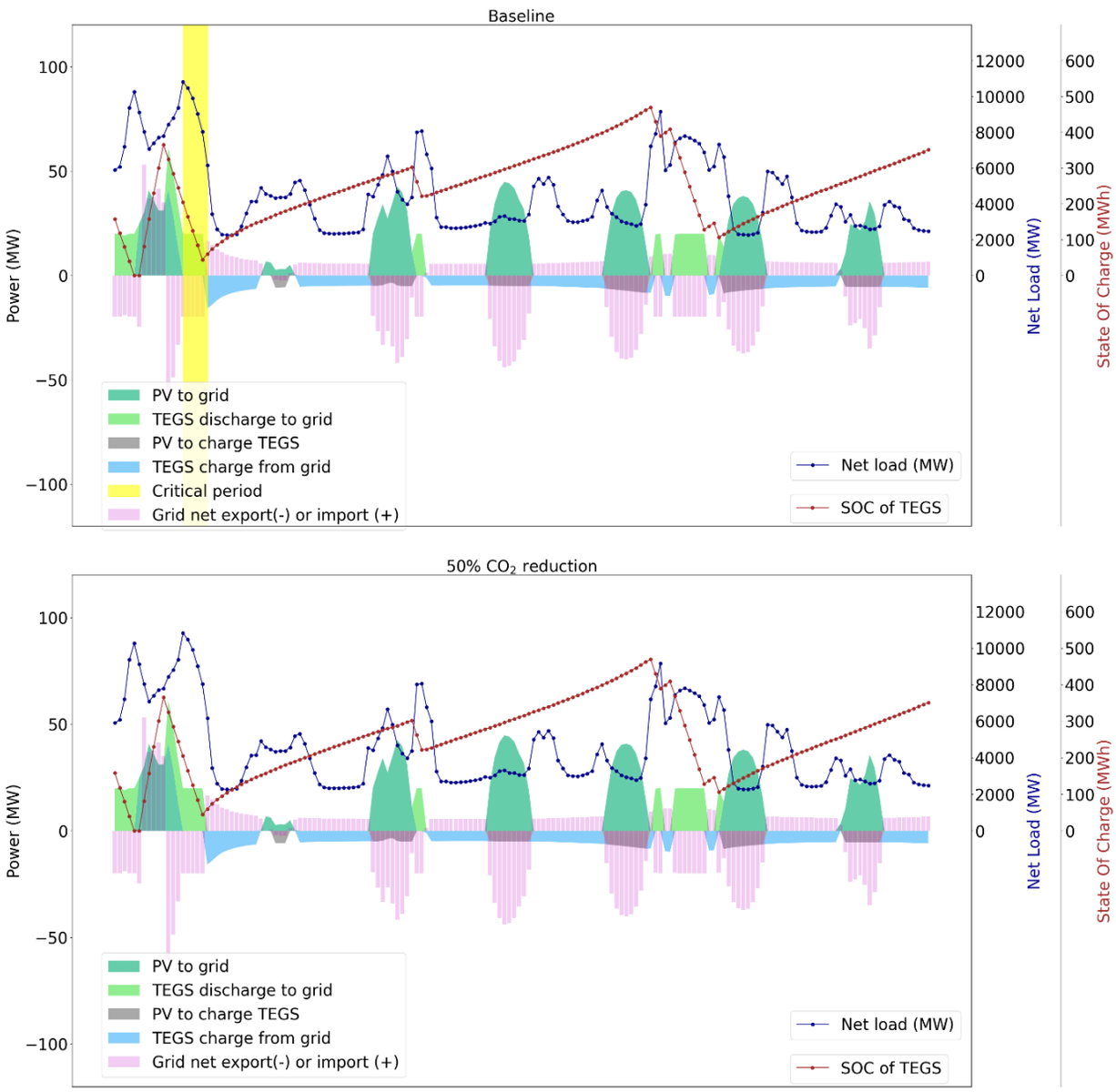


Figure 3: An example week with low electricity generation from the modeled 100 MW solar PV plant.

1.2 Electricity system cost

Since the CEM evaluates the cost-optimal portfolio of electricity generation technologies to serve a demand for electricity, it is of interest to investigate the resulting cost of the electricity system in the different CO_2 emission scenarios. Figure 4 illustrates the cost of the overall electricity system under different levels of CO_2 emission reductions.

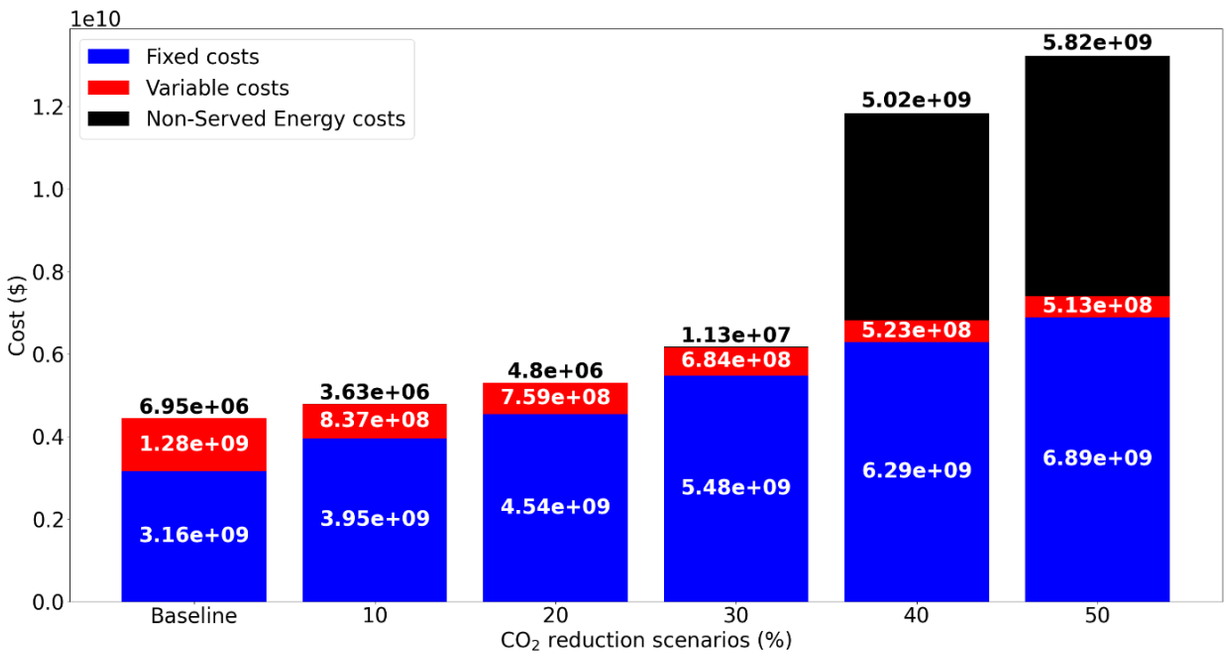


Figure 4: Electricity system cost under different CO_2 reduction scenarios.

Here the fixed cost is the total annualized investment and the fixed operation & maintenance (FOM) for the electric power system. The variable cost parameter is the total annual variable costs associated with all resources, including fossil fuel costs for thermal power plants. The cost associated with Non-Served Energy (NSE) is computed as the number of hours during the year when the demand is not met times the value of lost load (VOLL), which is set to be 2000 $/MWh for the modeled grid. The lowest cost is obtained when there are no constraints on CO_2 emission, as the grid can utilize a cheap NG to a high degree. Once the CO_2 emission is constrained, the cost of the grid increases as it cannot utilize cheap NG to the same degree anymore and must rely on other emission-free technologies. Noteworthy, the overall cost of the electricity system is adequate for the different emission reduction scenarios until it reaches a critical point where the CO_2 emission is required to be reduced by 40%. In this scenario, the grid should be heavily reliant on emission-free VRE technologies that have intermittent electricity generation, and the lack of balancing resources such as energy storage results that the grid is experiencing several periods when the electricity generation portfolio cannot serve the requested demand for electricity, resulting in a significantly increased cost related to NSE. Once the CO_2 emissions are required to be reduced by more than 50%, the objective function of the CEM model of serving the requested demand for electricity is no longer satisfied, and the existing portfolio of electricity generation and storage must be expanded to serve the electricity demand at all times. The modeled PV + TEGS system in this study is not sufficiently large to contribute to serving the demand at all times for the current portfolio of electricity generation when the CO2 emissions are heavily constrained. A suggested future work is to analyze how emerging storage technologies based on TES can contribute to cost-effectively reducing the cost of the overall

power system by investigating the amount of installed capacity that is required to enable full decarbonization of the grid.

1.3 Cost optimizing the engineering design of TEGS

As the TEGS unit stores energy at extremely high temperatures, the working temperature and heat loss of the TEGS system is two important design parameters when engineering the TEGS unit. The TEGS capital costs are divided into Cost Per Power (CPP) and Cost Per Energy (CPE). CPP ($/kW) shows the capital cost for charging and discharging, while the CPE($/kWh) is the capital cost of the stored energy. Figure 5 shows how the capital cost of the TEGS unit varies as a function of temperature and heat loss.

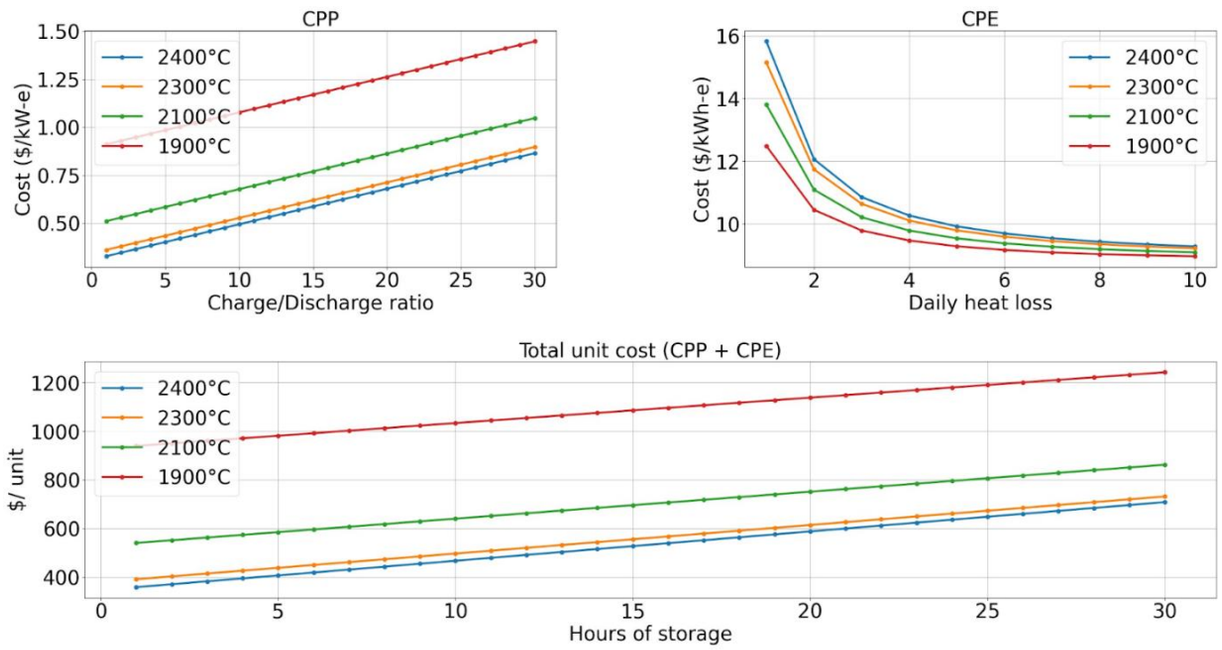


Figure 5: Different capital cost parameters of TEGS

The upper-right graph shows that the lowest temperature system (1900°C) has a lower CPE than the other designs. This is because of the lower need for system insulation when operating at lower temperatures. The CPE cost also reduces as a function of the daily heat loss as less insulation is required. However, the CPP is significantly higher when the temperature is lower. This is because of the lower power density (higher energy density at higher temperatures). Therefore, more Thermophotovoltaic equipment is required to discharge sufficient power to the electric power system. Therefore, the CPP was computed as a function of the charge/discharge ratio, which shows that the CPP increases linearly with respect to the charge/discharge ratio. Finally, the lowermost figure combines the CPE and CPP costs, showing that the TEGS unit is the cheapest at the highest possible temperatures. This motivated the choice of modeling a TEGS system at 2400°C.

However, the lowest capital cost does not necessarily mean this engineering configuration is the cheapest one when operating in a grid. Therefore, another modeling procedure was performed, where the TEGS engineering design was optimized with respect to minimizing the cost when operating in the grid. The most cost-effective design was found by minimizing the cost of the modeled electric power system compared to a baseline scenario where TEGS is replaced by a state-of-the-art Li-ion battery. Motivated by the results in Figure 5, two critical design parameters are modeled for the optimization: daily heat loss and the operating temperature. A matrix is created to

model various TEGS configurations with heat loss ranging from 1% to 10% and operating temperatures between 1900°C and 2400°C. The result is presented as a heat plot in Figure 6, where the red color shows the highest cost reduction, and the blue color shows the lowest cost reduction compared to a baseline.

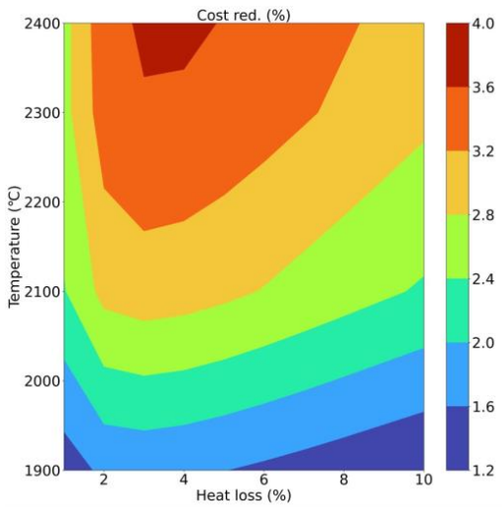


Figure 6: Optimized TEGS engineering design operating in an electric power system representing New England.

From the heat plot, the cost-optimum TEGS has an operating temperature of 2400°C and a daily heat loss of 1-3% compared to a baseline where TEGS is replaced by a Li-ion battery. In this case, the electric power system becomes approximately 3.5% - 4% cheaper than the baseline scenario.

1.4 Power availability from PV plus TEGS at Gigawatt scale

The capital cost projections of TEGS decrease with increasing size and are expected to achieve a capital cost below US$ 20 kWh^{-1} at the gigawatt scale [1]. Similarly, Apostoleris et al. [2] highlighted the record-low prices that could be achieved for PV plants at the gigawatt scale.

Motivated by the cost projections that can be achieved at the gigawatt scale for both the PV and TEGS, we replicate the modeling by considering a PV plant with an installed peak power capacity of 1 GW. In addition, the TEGS system had a storage capacity of 6 GWh. The total existing capacity in the modeled grid zone is the same as in the main manuscript (15 GW). Owing to the similarity of the results between the different storage sizes in the main manuscript, only one storage size is modeled in the Supplementary Information. The modeling method remains the same as in the main manuscript.

In Figure 7, the yearly PAF from the hypothetical PV + TEGS system as a function of the discharge power is shown. In addition, the PAF modeling results at megawatt and gigawatt scales were compared. The results show that the result for each scale is consistent because the PAF increases as the system is derated for both cases.

For the baseline scenario at the gigawatt scale, the PAF was higher than that at the megawatt scale. This is arguably due to the higher share of the gigawatt system when connected to the remaining grid, and thus, this system plays a more dominant role in supplying electricity demand with sufficient power. The PAF was approximately 55% and 65% when the discharging capacities were 100 MW and 1000 MW for the megawatt and gigawatt-scale systems, respectively.

Similar to the results at the megawatt scale (see the main manuscript), there was a large difference between the scenarios with and without CO_2 reduction constraints. The results are consistent for both the megawatt and gigawatt scales, whereas the PAF increases the more the system is derated. The dynamics of the electric power system change significantly when CO_2 emissions are reduced by 50%. For both scales considered, the increased retirement of NG makes the grid more dependent on the hypothetical PV + TEGS system. It is noteworthy that the megawatt-scale system has a slightly higher PAF than the gigawatt-scale system for the 50% CO_2 reduction scenario when the system is required to supply the grid with power between 500 MW and 1000 MW. This is due to changes in the electricity generation mix, and for the gigawatt scale, the least-cost CEM optimization results in that the PV + TEGS system has a few more periods during the year when the grid cannot receive the requested amount of power. However, the small discrepancy in the PAF between the megawatt and gigawatt scales shows that the modeling results are consistent at different scales.

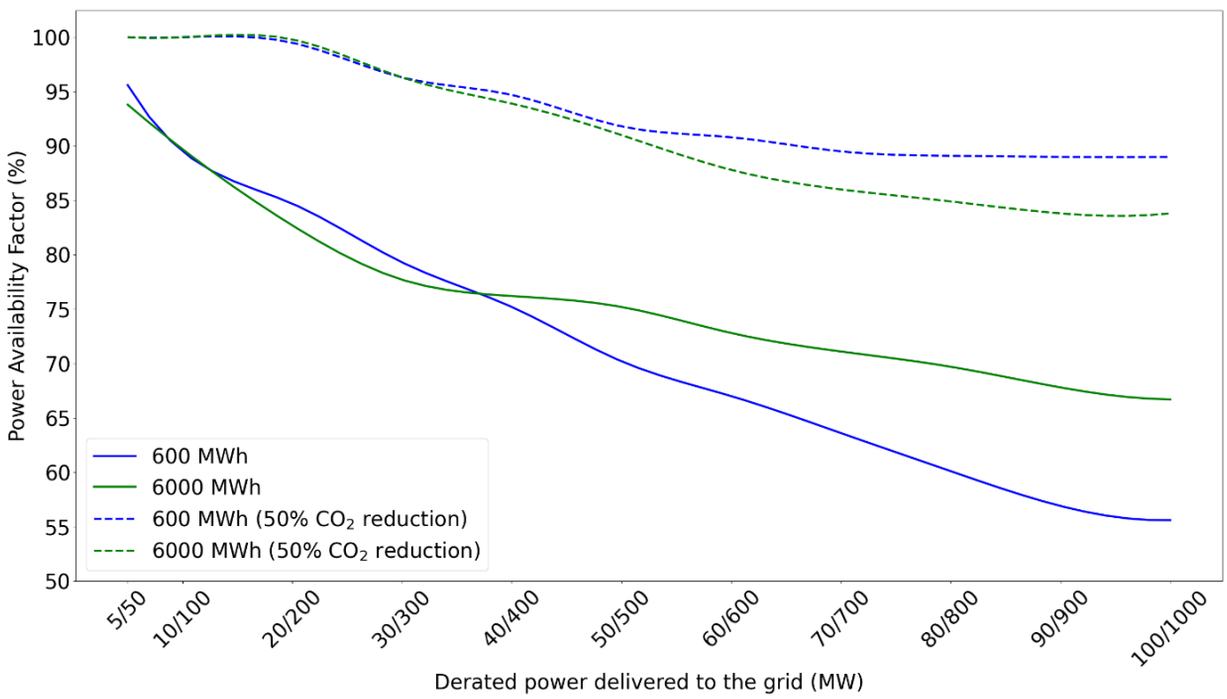


Figure 7: Percentage of time during the year when the hypothetical system can deliver the requested power to the grid. The percentage of time that the grid can receive the requested power increases by lowering the derated power. Here, the PV + TEGS system modeled at the megawatt scale delivers power in the interval between 5 MW and 100 MW, whereas the system at the gigawatt scale delivers power to the grid in the interval between 50 MW and 1000 MW.

The differences between the emission scenarios are shown in Figure 8. Here, hourly operation with a 6 GWh storage unit and 200 MW discharge capacity is shown during a typical winter week in January. The uppermost and lowermost figures show how the system operates in scenarios with and without constraining the CO_2 emissions.

From both graphs, TEGS is used frequently to discharge power to the grid whenever there is low electricity generation from the solar PV. In addition, the TEGS charges power from the grid whenever there is a drop in demand to increase the SOC.

For the gigawatt scale, as for the megawatt scale in the main manuscript, there are one critical period where the PV + TEGS system cannot deliver the requested electricity to the grid because the SOC is already zero.

Considering the 50% CO_2 reduction scenario, the system interacts differently with the grid, and the CEM optimizes the hypothetical system to have more energy available at more times because the grid must be less dependent on the electricity generation from NG.

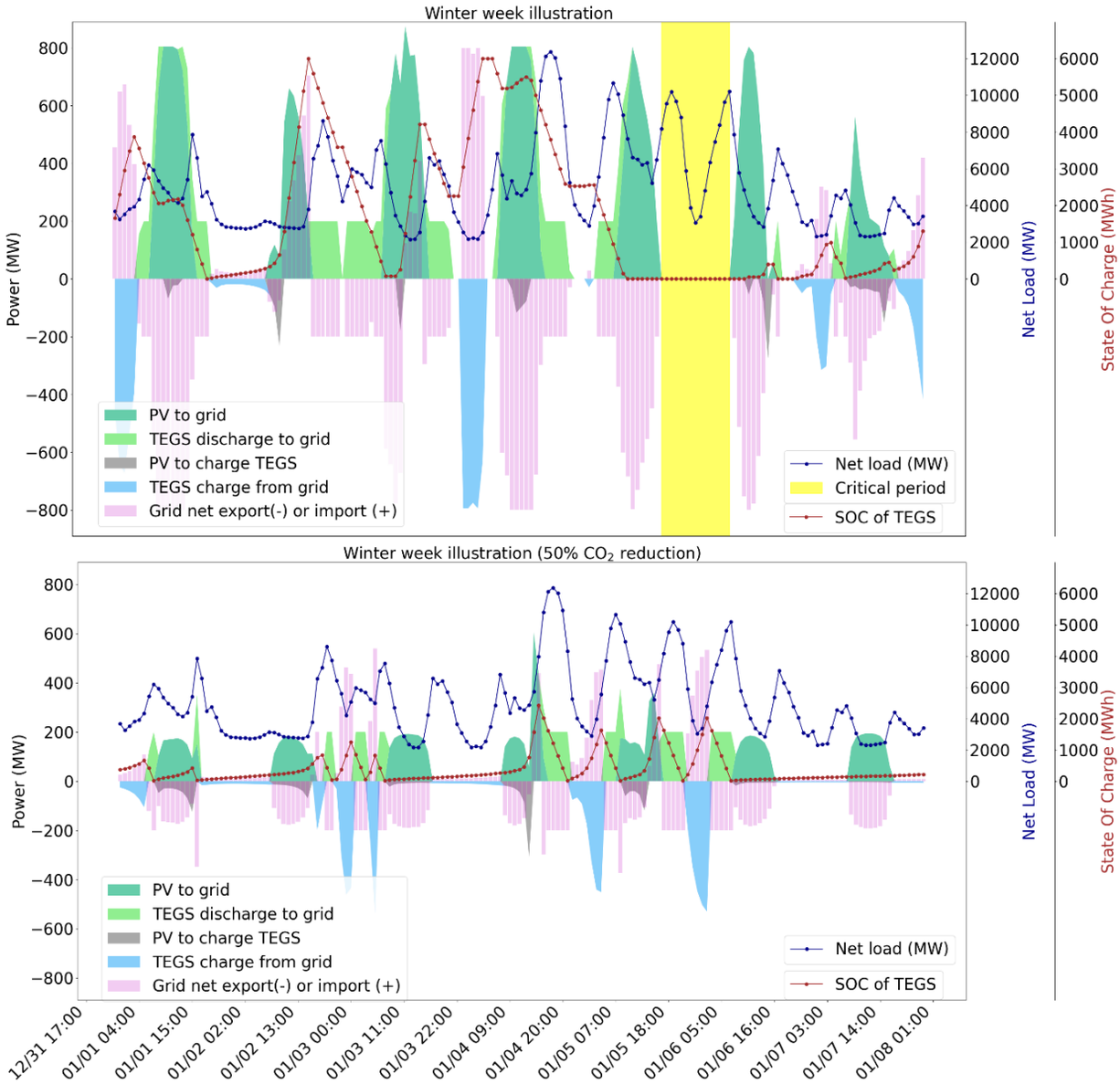


Figure 8: Winter week illustrations for a 1000 MW PV plant plus a 6000 GWh TEGS system that can discharge 200 MW. Baseline case without constraining CO_2 emissions (uppermost graph) and future scenario with lower emissions (lowermost graph). Both cases show no critical period during which the hypothetical system cannot deliver the requested power to the grid because the SOC of TEGS is already zero.

In addition to the winter-week example shown in Figure 8, a typical summer week is illustrated in Figure 9. In the summer, the higher solar availability ensures that the system can deliver most of the required power to the grid from the PV system, and TEGS is used less frequently.

Compared to the winter week example, there are now several periods the hypothetical system cannot deliver the requested power to the grid because the SOC of TEGS is already zero. This is due to the lack of CO_2 constraints which ensure that the grid is mainly dependent on NG. Similar to the megawatt-scale simulation in the main manuscript, the TEGS system is only used to provide additional peak power when the demand suddenly increases for the baseline scenario.

Under the 50% CO_2 reduction scenario, the grid is again more dependent on the discharge power from the TEGS unit whenever there is no electricity generation from the PV plant. In this scenario, instead of providing peak power to the grid, the PV system is used to charge the TEGS unit to a higher degree to ensure that the SOC is never zero. Thus, the derated power can always be discharged to the grid when the PV system does not generate electricity.

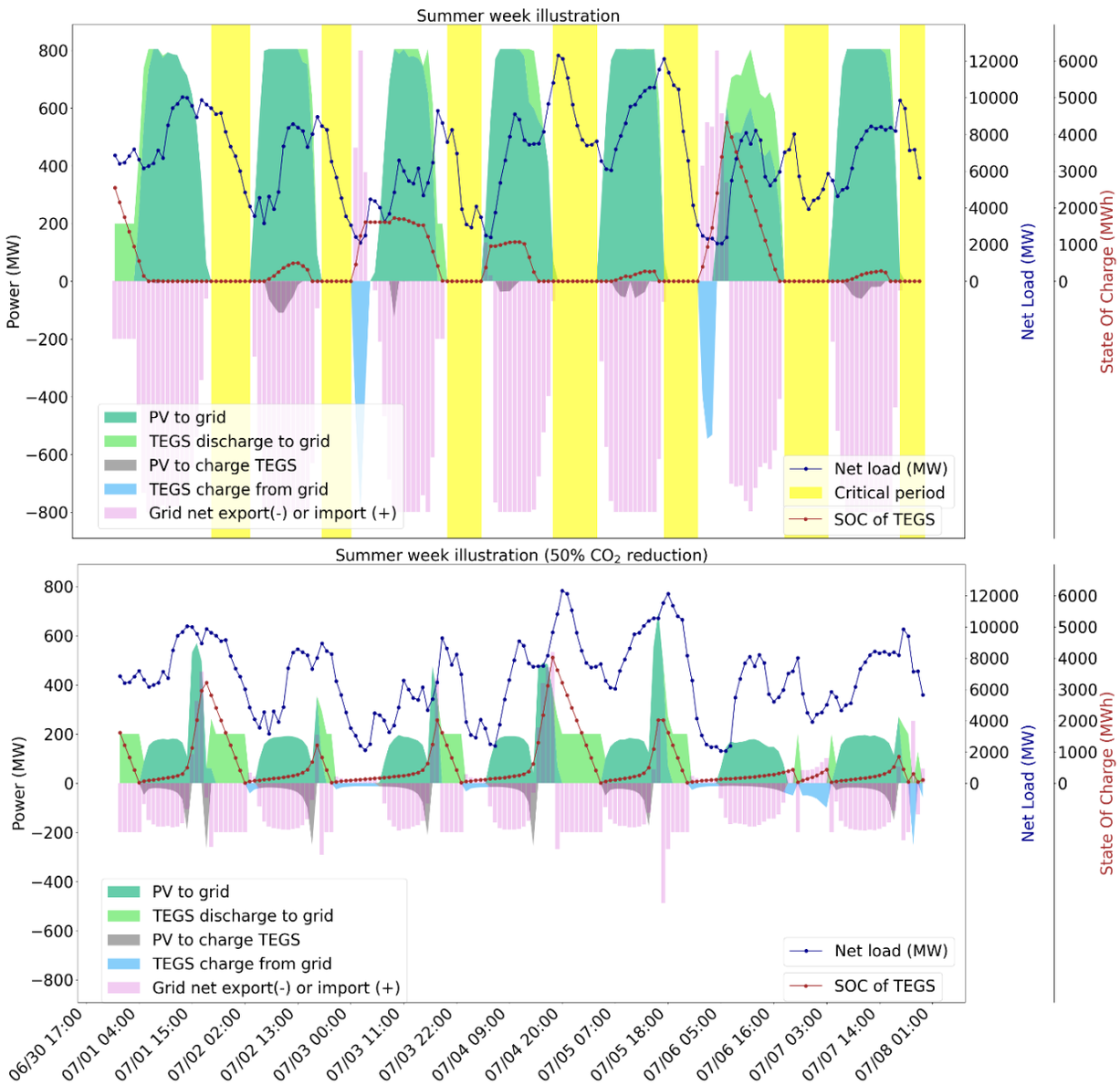


Figure 9: Illustrations of a typical summer week for a 1000 MW PV plus 6000 MWh TEGS system that can discharge 200 MW. There are several periods during which the system cannot deliver the requested power to the grid because the SOC is already zero, whereas such periods never occur in the CO2 reduction scenario.

References

[1] C. C. Kelsall, K. Buznitsky, and A. Henry, "Technoeconomic analysis of thermal energy grid storage using graphite and tin," *arXiv preprint arXiv:2106.07624,* 2021.

[2] H. Apostoleris, S. Sgouridis, M. Stefancich, and M. Chiesa, "Evaluating the factors that led to low-priced solar electricity projects in the Middle East," *Nature Energy,* vol. 3, no. 12, pp. 1109-1114, 2018/12/01 2018, doi: 10.1038/s41560-018-0256-3.

10

Paper V

COST-EFFECTIVE THERMAL ENERGY GRID STORAGE FOR DECARBONIZING ELECTRIC POWER SYSTEMS

Odin Foldvik Eikeland[1,3,*], **Ruaridh Macdonald**[2,*], **Harry Apostoleris**[4], **Shomik Verma**[1], **Kyle Buznitsky**[1], **Matteo Chiesa**[1,3,5,**], **& Asegun Henry**[1,**]

[1]Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA.
[2]MIT Energy Initiative, Massachusetts Institute of Technology, Cambridge, MA, USA.
[3]Department of Physics and Technology, UiT – the Arctic University of Norway, 9037 Tromsø, Norway.
[4]Dubai Electricity & Water Authority
[5]Laboratory for Energy and NanoScience (LENS), Khalifa University of Science and Technology, Masdar.

SUMMARY

As variable renewable energy sources comprise a growing share of electricity generation, energy storage is becoming increasingly critical for balancing generation and demand. The high capital cost of electrochemical batteries limits the integration of renewable energy that can be incorporated into electric power systems without substantial cost increases. This study models a thermal energy grid storage unit ("thermal battery") called TEGS with estimated capital cost sufficiently low to enable large-scale deployment in electric power systems. The findings indicate that optimizing the engineering design of TEGS can potentially reduce the cost of decarbonization by approximately 4% compared to a baseline scenario without TEGS. Achieving such savings requires constant charge and discharge capacities and avoiding negative state of charge. We further explore how TEGS enhances grid resiliency in heavily decarbonized grids by modeling the grid's resilience under 22 different weather years. The results demonstrate that TEGS significantly improves electric power grid resilience.

***K*eywords** Thermal Energy Grid Storage · Electric Power System Modeling · Decarbonizing Electric Grids

[*]These authors contributed equally to this work
[**]Corresponding authors: mchiesa@mit.edu and ase@mit.edu

1 Introduction

To decarbonize electric power systems, variable renewable energy (VRE) resources like wind and solar energy must be increasingly utilized [1, 2, 3]. The inability to directly modulate the electricity supply to match energy demand due to VRE sources' weather dependency makes it more challenging to operate electric power systems. To compensate for this variability, energy storage is becoming an increasingly important component of future energy systems. Energy storage can store surplus electricity during high-generation and low-demand periods. During periods of low generation from VRE sources, the stored electricity can be discharged to ensure adequate supply during periods of high electrical demand [4, 5].

The growing prevalence of VRE technologies in electric power systems has increased the demand for low-cost storage over a range of timescales, spanning seconds to days and even seasonal storage [6, 7]. A well-established method for large-scale energy storage, pumped hydropower storage (PHS) is restricted to certain geographic areas. Concentrated solar power with thermal energy storage (CSP-TES), has been seen as promising, but projects worldwide have experienced delays, cost overruns, and mechanical problems, and the interest has recently fallen off[8, 9]. Lithium-ion batteries, whose production has scaled up dramatically with the widespread adoption of electric vehicles, are currently not cost-effective for the overnight and multiday storage systems needed to fully decarbonize the grid due to capital costs of between US$150 and US$300 kWh^{-1} [10, 4, 5, 11, 12]. Another type of electrochemical storage is Redox-flow batteries, which are proposed as a promising technology for long-term storage. Still, the relatively high capital cost of 170 kWh^{-1} has decreased the interest for this storage technology and currently accounts for less than 1.5% of energy storage systems under development [12]. Studies suggest that capital costs must be reduced to \$3-30 kWh^{-1} to achieve cost-effective multiday storage [5, 13].

A storage approach based on Thermal Energy Storage (TES) has demonstrated promising potential for obtaining sufficiently low capital costs for the multiday storage regime. The Thermal Energy Grid Storage (TEGS) technology described in [14, 15] uses thermophotovoltaic (TPV) cells utilized to convert the energy that is stored as heat in graphite storage blocks to electricity as needed [16].

From the second law of thermodynamics, converting heat to electricity results in significant efficiency penalties. However, storing energy as heat instead of electrochemically can be vastly cheaper. Thus the round-trip efficiency (RTE) penalty compared to electrochemical batteries ($\sim$ 90%) can potentially be a worthwhile tradeoff [14]. Heat is stored at very high temperatures ($\sim$2400 °C) to enhance the thermal-to-electric conversion efficiency. A world-record high conversion efficiency of 41 % utilizing TPV was demonstrated by the authors in a recent study [17], which predicted a conversion efficiency of 50 % in the future. As a result, at gigawatt scales, this technology can achieve a projected cost of less than US\$ 20 kWh^{-1}.

TEGS has the unique ability to decouple the charge and discharge capacities of the storage unit since energy is first stored as heat in graphite blocks and then converted to electricity using TPV. In contrast to what is needed for discharging, the TEGS can now charge (i.e., store heat) at a significantly higher capacity. The advantage of having such a property is that a significant amount of energy can be quickly charged during generation surpluses and discharged over a more extended period to fulfill the electricity load when demand exceeds supply.

The existing literature includes several studies that have employed different approaches to evaluating the storage requirements (capital cost and storage duration) to enable cost-effective decarbonization of the electric power system [7, 12, 18, 4, 19, 20, 21, 22, 23, 5, 11, 24, 25].

Sepulveda et al. [7] addressed the cost requirement of storage to replace firm technologies by analyzing projected cost assumptions of different storage technologies. The findings show that the storage cost must be below U.S. \$20 kWh^{-1} to reduce the electricity cost by $\geq$ 10% compared to an electricity system powered by firm low-carbon generation.

A comprehensive analysis of the value of energy storage in decarbonized grids with a high share of VRE technologies[12] found that a cost-optimal solution is to deploy multiple storage technologies. The technologies with the lowest capital cost of energy storage capacity are generally best suited for long-term storage in decarbonized power systems.

Braff et al. [4] addressed the value of using energy storage to balance the intermittent generation from renewables by proposing a framework to compare storage technologies and then set cost improvement targets. The authors concluded that storage technologies add value to solar and wind energy, but cost reduction is required to reach overall profitability. The authors also find that the optimal cost trajectory is relatively location invariant, which can inform broad industry and government technology development strategies.

The authors in [21] investigate the role of using firm low-carbon resources (nuclear, reservoir hydro, geothermal, bioenergy, and fossil plants capturing CO_2) in decarbonizing power generation in combination with VRE, battery energy

storage, demand flexibility, and long-distance transmission. They evaluated scenarios consider varying CO_2 limits, technological uncertainties, geographical differences in electricity demand, and renewable resources. The findings show that including firm low-carbon resources reduces the electricity cost by 10% - 62% compared to cases where the electricity system relies only on wind, solar, and batteries.

In [5], the authors studied the characteristics that most impact renewable electricity costs, including cost features of proposed storage technologies. To consider VRE resource fluctuations, they analyzed 20 years of data to capture significant, infrequent events affecting storage requirements. The findings show that the stored energy capacity cost must be below U.S. $20/kWh to cost-effectively meet a baseload demand for electricity.

The authors in [11] assessed the holistic value of utilizing energy storage in future decarbonized grids with increasing wind and solar generation. The authors identified significant sources of storage value and their dynamics under different systems settings and at increasing storage, wind and solar penetration levels. The authors find that increasing VRE penetration from 40% to 60% improves storage value, but only enough to make storage capacity up to 4% of peak demand cost-effective at current Lithium-ion capital cost. The authors used a future capital cost of U.S. $ 150/kWh for 4-hour duration storage. Increasing the storage duration increases the storage value, but the high capital cost of Lithium-ion often exceeds the storage value.

The former literature agrees that the energy storage capex must reach U.S. \$20 kWh^{-1} in order to suitable for long-duration applications. However, previous work has yet to model an emerging storage concept based on TES and optimize the engineering design to match specific market conditions to reduce the overall cost to enable cost-effective grid decarbonization.

In this study, we propose a modeling framework for cost-optimizing the engineering design of TEGS to operate in electric power systems. Using a Capacity Expansion Model (CEM) [26], we model two hypothetical grid regions in the U.S. Northeast (North) and Texas (South) that has different climate conditions to investigate the optimum design under various market conditions. We find the optimum engineering design of TEGS under different scenarios where the CO_2 emissions are constrained. In addition, by comparing to a baseline scenario with no TEGS, we analyze how introducing such technologies can reduce the cost of decarbonization under different CO_2 reduction scenarios. We also assess the marginal price of adding additional TEGS units into the electric power system to identify the storage size that mostly contributes to reducing the cost of the electric power system.

Despite much literature addressing the storage requirements to decarbonize electric power systems cost-effectively, fewer studies combine such studies with addressing how storage contributes to maintaining resiliency. Maintaining resiliency in decarbonized electric power systems is becoming increasingly difficult due to the intermittent and uncontrollable electricity generation from VRE sources. Therefore, we explore how optimizing storage design can improve grid resiliency by computing the annualized non-served energy (NSE) cost. The NSE cost is computed as the number of hours during one year when the electricity demand is unmet. To assess the grid resiliency in decarbonized electric power systems using TEGS, we compare the grid resiliency with a baseline scenario with no TEGS available.

The main contribution of this study is: Rather than modeling the general storage requirements in capital cost of energy storage capacity, we optimize the engineering design of a TEGS unit to investigate the configuration that provides the most cost-efficient power system decarbonization compared to a baseline scenario. The engineering parameters of the TEGS unit that are optimized are; the operating temperature, daily heat loss, and the ratio of maximum charging and discharging capacity. In addition, we compute the annualized NSE cost over 22 different weather years to indicate how optimized storage units can affect grid resiliency in future electric power systems that are heavily dependent on VRE technologies.

The remainder of this paper is organized as follows. Section 2 presents the optimized design of the storage unit under different scenarios. This section also highlights which technologies are used to provide electricity to customers under the different CO_2 reduction scenarios. We also discuss how introducing TEGS can affect grid heterogeneity, indicating the expected grid resiliency. Finally, we examine whether optimized storage technologies can enable efficient stepwise decarbonization of the electric power system towards the year 2050. Finally, the conclusions are provided in the last part of Section 2. Section 3 details the experimental procedures for presenting the electric power system case studies. In addition, a description of the TEGS engineering optimization method is described.

2 Results and discussions

2.1 Optimized thermal battery design to reduce overall power system cost

Fig. 1 shows the resulting optimized TEGS under different CO_2 scenarios for the New England power system. The red color in the heatmap indicates the highest values, while the blue colors indicate the lowest values.

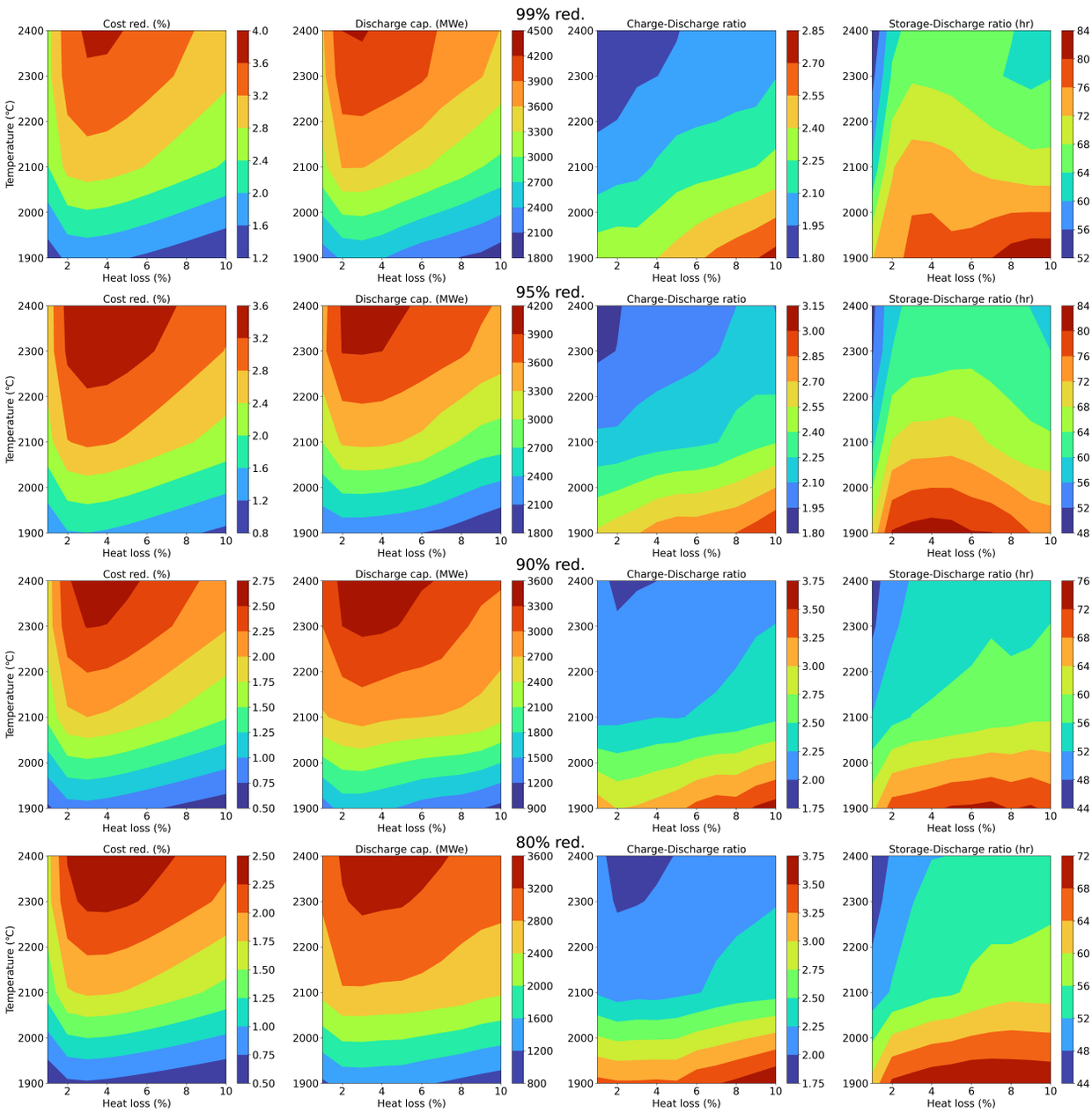


Figure 1: Optimized TEGS configurations under different CO_2 reduction scenarios

The leftmost contour plot in the uppermost row (99% CO_2 reduction scenario) shows the cost reduction compared with the baseline scenario (i.e., TEGS excluded as an available technology) for each storage configuration. The optimum configuration results in a 3.5% cost reduction compared to the baseline case and is obtained when the TEGS has an operating temperature of 2400°C and a daily heat loss of approximately 3%. The second contour plot to the right shows that approximately 4000 MW of discharge capacity is required to reduce the overall grid cost by 3.5%. The

third contour plot to the right computed the ratio between the discharge and charge capacities. Interestingly, it is clear that the optimum configuration requires TEGS to charge at a higher capacity than discharging. To obtain the highest possible cost reduction compared with the baseline, the storage unit should have a charging capacity that is approximately 1.5 times higher than the discharging capacity. This is because when the electricity demand is low, the storage facility should be able to charge cheap electricity fast and discharge the electricity over a longer period when the electricity demand is high. The rightmost plot shows that approximately 50 hours of storage is required to obtain the cost reduction.

The second to fourth rows show the cost reductions for the 95%, 90%, and 80% CO_2 reduction scenarios. The results are similar for all scenarios, where the most cost-optimized TEGS is obtained with a working temperature of 2400°C and daily heat loss of approximately 3%. The supplementary material shows that the Electric Reliability Council of Texas (ERCOT) region has similar trends. Fig.1 in the Supplementary material shows that the optimum TEGS has a working temperature of 2400°C and daily heat loss of 3%. Similarly to the New England power grid, the cost reduction compared to the baseline is approximately 3%. As the ERCOT grid in Texas covers a significantly larger area, approximately 25000 MW of charging capacity is required for the 99% CO_2 reduction scenario. The optimum charge-discharge capacity is approximately 3.75. This is significantly higher in the New England grid, where the storage unit should charge with a maximum capacity that is 1.5 times higher than the maximum discharging capacity. The vast difference in charging requirements for the two different power systems is because the share of solar energy in the Texas grid is significantly higher than in the New England grid, where wind power dominates. In a solar-heavy power system, the periodic generation profile of the solar PV systems gives excess electricity during mid-day when the sun is high but no electricity at all during the night. This requires storage systems to be able to charge at a high rate when there is excess electricity during mid-day. Then the stored energy can thereafter be discharged over a longer period when solar generation is low or zero.

The optimal charge-discharge ratio can vary depending on variables such as the availability of renewable energy sources, electricity demand patterns, and the market pricing structure. TEGS design can address this issue of fluctuating market conditions by altering the component sizes of the system. By adjusting the relative sizes of heating elements, thermal storage blocks, and TPV units, the TEGS system can be optimized to meet the unique requirements of a particular region or market

As discussed in the Experimental Procedures section 3.2, the system's effective charge and discharge capacity varies as a function of State of Charge (SOC). In addition, due to the extremely high temperatures when the TEGS unit is fully discharged, it can still experience heat losses when SOC = 0. Consequently, if not recharged again directly after SOC = 0, the TEGS unit can obtain a negative SOC since the system still has heat losses when fully discharged. Therefore, we performed another modeling procedure with varying charge and discharge profiles and the possibility of obtaining a negative SOC if the TEGS unit is fully discharged and is not directly recharged again. . The results are provided in Fig. 2 and show the optimum engineering design of a TEGS unit considering the effective charge and discharge capacity changes as a function of the SOC and the possibility of obtaining negative SOC.

The result shows a similar optimum design for the TEGS unit as in Fig. 1, but with an optimum heat loss of 2% instead of 3%. This is due to the constraint concerning the possibility of obtaining a negative SOC when the TEGS unit is not immediately recharged after full discharge. As discussed in the Experimental Procedure section 3.2, the charge-discharge and storage-discharge ratios must be fixed to implement the constraints of effective charge and discharge capacities into the CEM. Therefore, the charge-discharge and the storage-discharge rations are the same as in Fig. 1.

However, the cost reduction compared to the baseline scenario is significantly lower for all CO_2 reduction scenarios, with a cost reduction of 0.5% for the 99% reduction scenario. The installed discharge capacity is 1750 MW for the optimum configuration compared to 4500 MW for the optimum configuration in Fig. 1. The supplementary material shows that the EROT region has similar trends.

Such a large impact on cost reductions highlights the importance of researching solutions to maintain a constant capacity during charge and discharge, and avoid periods with negative SOC. However, the results presented in Fig. 2 are considered a worst-case scenario as the system's true charge and discharge profiles are in between the two results presented in Fig. 1 and Fig. 2. The TEGS units' true charge and discharge capacities have non-linear profiles, and ongoing research is to identify these profiles which can be implemented into the CEM. In addition, obtaining solutions to avoid periods with negative SOC is suggested as an important study for thermal energy storage technologies. In the following, as we model future system assumptions, we assume that R&D has enabled the TEGS unit to have constant charging and discharging capacities with no periods where SOC is negative (i.e., the results in Fig. 1). Therefore, the following results present the optimum TEGS system with constant discharge and charge capacities.

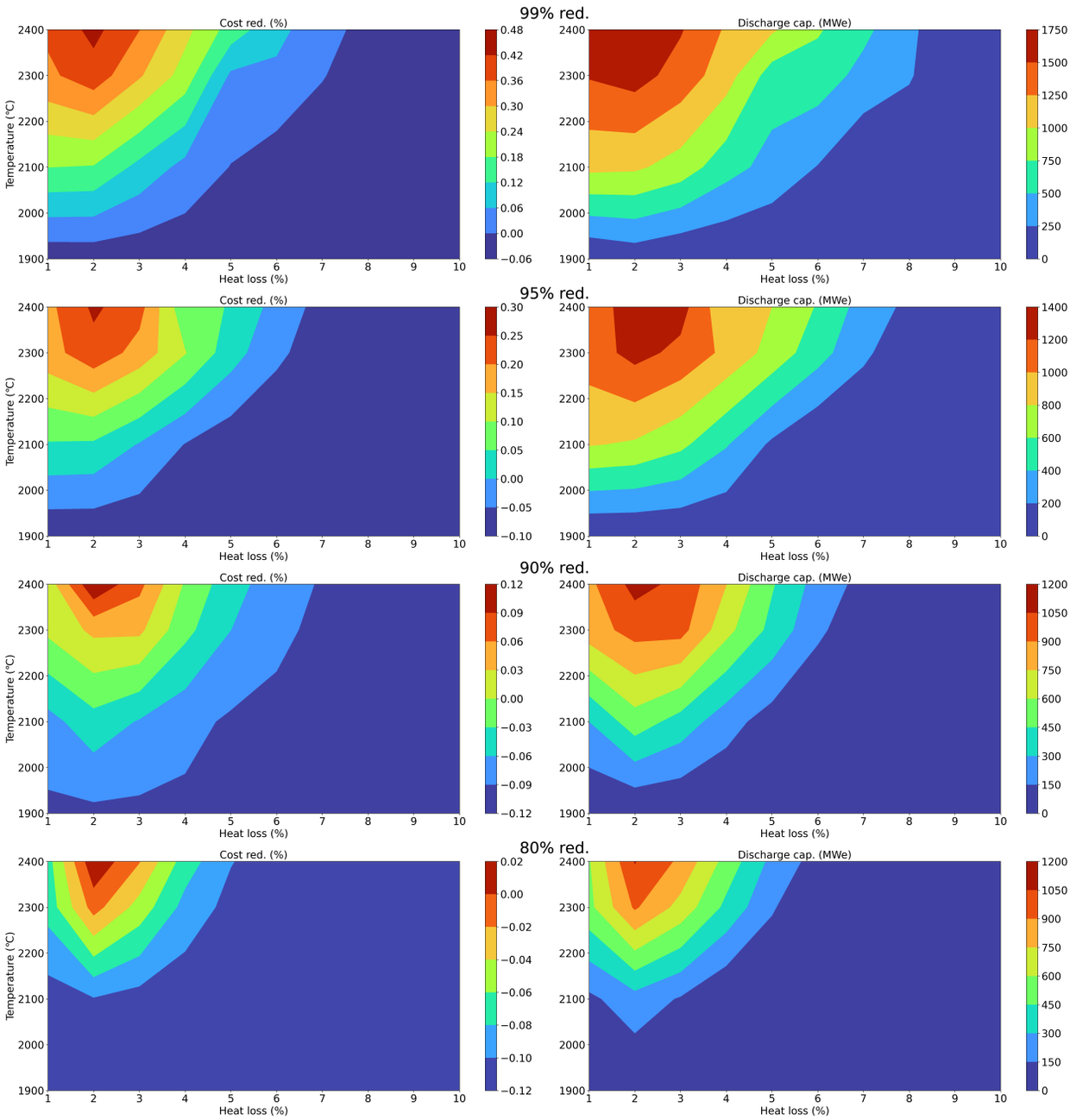


Figure 2: Optimized TEGS configuration under different CO_2 reduction scenarios considering the effective discharge and charge capacities as a function of SOC and the possibility of obtaining negative SOC.

2.2 Installed capacity under different scenarios

Investigating how the mix of electricity generation technologies changes for the different CO_2 reduction scenarios is interesting. Fig. 3 shows the optimum portfolio of electricity generation technologies in the New England and ERCOT power systems under different decarbonization scenarios using the engineering-optimized TEGS unit.

The cyan line indicates the total installed capacities for all decarbonization scenarios. More capacity must be installed in the New England and ERCOT grid to meet the demand when the CO_2 reductions are higher. In heavily decarbonized grids, the high share of non-dispatchable renewables (i.e., solar and wind power) requires installing a large capacity to cover peak demand. Consequently, during periods with lower demand, the electric power system has an oversupply of

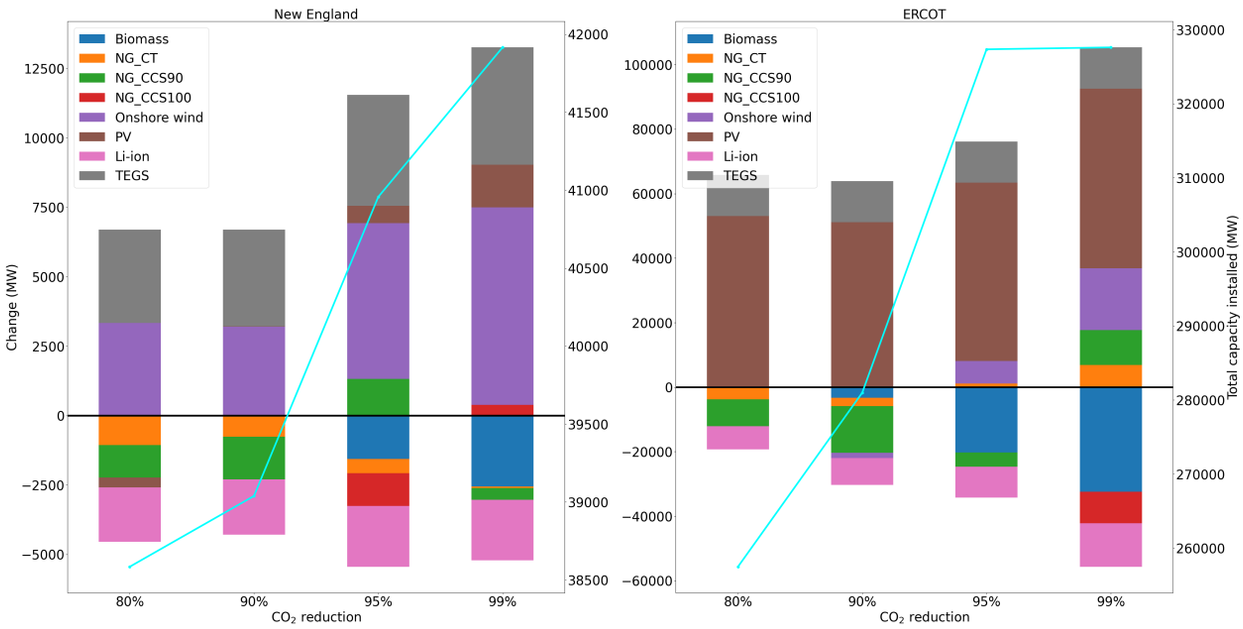


Figure 3: Change of installed capacity in the New England and ERCOT power system at different decarbonization scenarios

electric generation. On the contrary, dispatchable fossil-fuel-based technologies can be sized to cover the electricity demand in peak periods and periods with a lower need for electricity, thus reducing the need for net installed capacity in the electric power system.

All technologies with a negative value indicate a reduced installed capacity compared to the baseline case where TEGS is not included. Clearly, the installed capacity of Li-ion batteries has significantly decreased as TEGS is replacing Li-ion. For the 99% reduction scenario, TEGS reduces the need to utilize the Biomass and Natural Gas (NG) Combined Cycle with Carbon Capture and Storage (CCS). On the other hand, introducing TEGS allows for a significantly increased amount of onshore wind power. In addition, TEGS allows for more installation of PV compared to the baseline case. In the ERCOT grid, less wind power is installed, but a large amount of solar power is installed due to the large solar availability compared to the New England grid region.

When the CO_2 emissions are less constrained, there is an increased amount of NG with 90% CCS, while there is a decreased amount of NG with 100% CCS in the New England grid. This is because when the emissions are less constrained, the optimized grid utilizes NG with 90% CCS instead of 100% CCS as this technology has a lower capital cost. In the ERCOT grid, the installed capacity NG with 90% CCS is also reduced when the CO_2 reductions are less constrained.

In all reduction scenarios, the installed capacity of onshore wind is similar in the New England grid. Similarly, the installed solar energy capacity is similar in the ERCOT grid for all reduction scenarios. This indicates that wind power and solar energy are cost-efficient technologies regardless of CO_2 emissions constraints. Overall, introducing TEGS provides a higher share of renewables and a lower share of fossil-fuel technologies in the New England and ERCOT grid.

2.3 Shadow prices

In addition to identifying the engineering-optimized TEGS and the portfolio of supply technologies under different decarbonization scenarios, this section presents the shadow prices of the optimum TEGS design under the 99% reduction scenario in the New England grid. The Supplementary material shows the shadow prices for the ERCOT grid. The objective of computing the shadow price is to investigate the marginal value of adding additional TEGS capacity into the power system. Fig. 4 shows the evolution of the marginal price for the optimum TEGS design with different discharge capacities.

Since the objective function of the CEM is to minimize the cost of the generation portfolio to meet electricity demand, negative shadow prices will decrease the objective function and thus further reduce the cost of the electric power system.

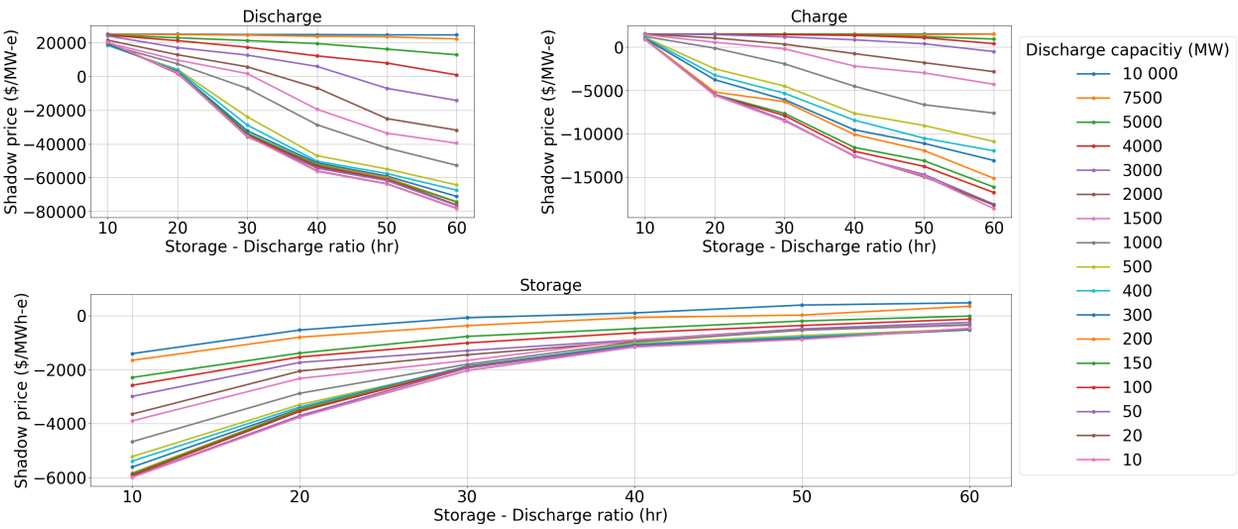


Figure 4: Shadow prices of TEGS

In Fig. 4, the upper left graph shows the shadow price of discharging as a function of the storage-discharge ratio. All discharge capacities have a positive shadow price below 20 hours of storage, indicating that adding another TEGS unit with less than 20 hours of storage does not reduce the electric power system costs. Interestingly, for large TEGS units with a discharge capacity above 5000 MW, adding another TEGS unit does not reduce the cost of the electric power system. This aligns with the results in Fig. 1, which shows that the optimum size of the TEGS unit should have a discharge capacity between 4200 MW and 4500 MW. For smaller TEGS units, the shadow price decreases significantly after 20 hours of storage, indicating that more units of such sizes positively impact the objective function of the CEM to reduce the overall electric power system cost. The shadow prices of charging are indicated in the upper right graph. Similarly, the charging shadow price reduces for longer storage durations. The lower graph shows the shadow prices of adding storage to the electric power system. Interestingly, here all storage sizes (discharge capacity) show negative values, indicating that adding more storage contributes to reducing the cost of the electric power system. The increase in shadow price indicates the cost reduction of adding another storage unit to the electric power system decreases with a larger storage capacity (i.e., hours of storage). Overall, the shadow prices show that adding TEGS to the electric power system effectively reduces the cost of the electric power system and gives insight into the impact of adding additional TEGS units with respect to reducing the cost of the electric power system.

2.4 Reliability of decarbonized electricity grids

In this section, we investigate the reliability of the resulting optimized grid under the 99% CO_2 reduction scenario with the optimized TEGS (2400°C and 3% daily losses). In particular, we model the resulting optimized grid as a brownfield model (i.e., we start from the resource mix obtained in the 99% CO_2 reduction scenario in Fig. 3) and model the resiliency of the grid when it is exposed to different weather conditions. This allows us to investigate how resilient a grid that is heavily dependent on VRE sources is under different weather conditions and whether storage units increase the grid's resiliency. We compute the resulting annualized non-served energy (NSE) cost to measure resiliency. The NSE cost is computed as the number of hours during the year when the demand is not met times the value of lost load (VOLL), estimated to be 2000 $/MWh for the New England power system. In summary, the higher the NSE cost, the less reliable the grid is, as there are more incidents during the year where the grid cannot meet the demand for electricity. To test the grid reliability under different VRE availability scenarios, we model 22 weather years from the year 2000 to 2021. Due to the lack of data availability for the ERCOT region, the reliability from the year 2000 to 2021 is computed for the New England region only. Fig. 5 shows the NSE cost for each year for the baseline scenario and the scenario with the optimum TEGS configuration under the 99% CO_2 reduction scenario.

In Fig. 5, it is clear that the grid has the lowest NSE cost when the wind and solar capacity factor is high (i.e., high availability of solar and wind power). The NSE cost increases when VRE sources have low availability (2003 and 2008). This is because when the CO_2 emissions are heavily constrained, the high dependency on VRE technologies makes the grid vulnerable to changes in electricity generation from renewable technologies such as solar and wind power. Interestingly, during the 21 weather years, the solar and wind capacity factor has had minor changes, indicating that the availability of solar and wind power resources has been consistent for decades. It is also clear that TEGS vastly

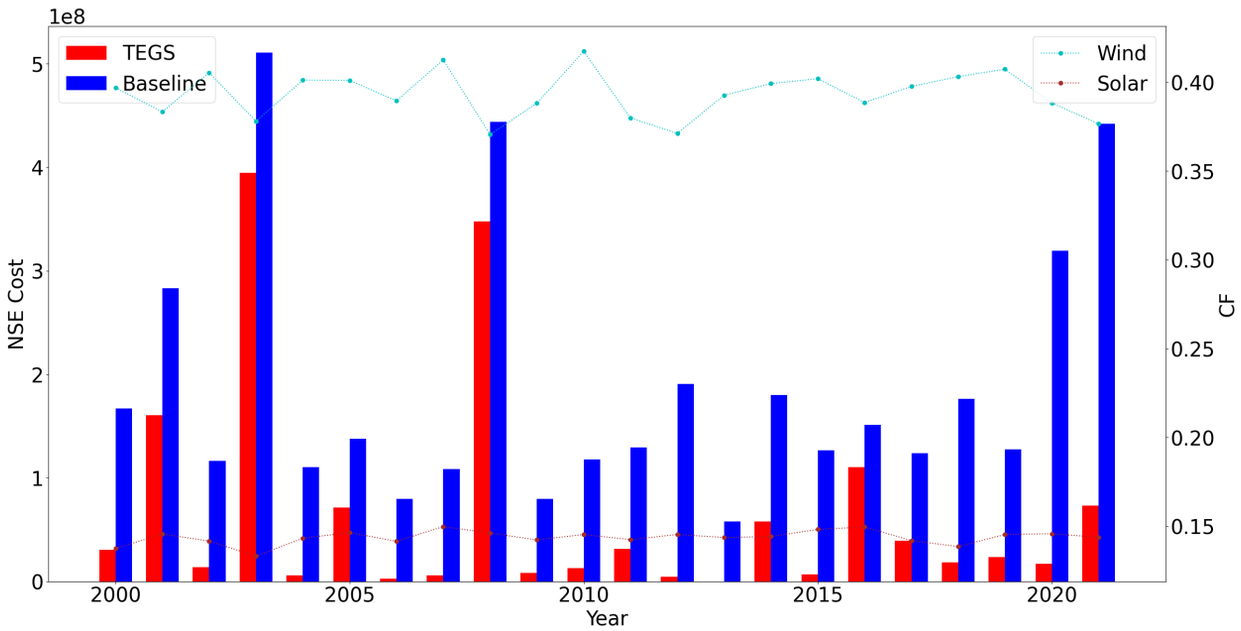


Figure 5: Cost of non-served energy for weather years with different VRE availability

reduces the cost of NSE compared to the baseline case due to the larger amount of available storage technologies that can serve the grid when there is a lack of solar and wind availability.

This has important implications for decarbonized electric power systems that are heavily dependent on non-dispatchable electric generation resources such as solar and wind power. Here it is shown that TEGS can cost-effectively store energy over longer durations and discharge the required power to cover a specific demand for electricity when there is a lack of availability from renewables (i.e., due to adverse weather conditions), thus ensuring a more reliable decarbonized electricity system.

2.5 Conclusions

In this study, we analyzed how storage design can be optimized to reduce the overall cost and enable full decarbonization of electricity systems. Using a capacity expansion model, we modeled two electric power systems in North America, representing the New England grid in Massachusetts and the Electric Reliability Council of Texas (ERCOT) grid in Texas. Modeling two different geographical regions allows for investigating how the requirements of energy storage change as a function of different electricity market conditions.

Because of the high capital cost of electrochemical batteries, a TES technology with a projected capital cost that fulfills the requirements (< US$ 20 kWh^{-1}) to enable full decarbonization of the grid was considered. Different storage design parameters were optimized to identify the technology that resulted in the most cost-efficient decarbonized electricity system. The optimized design parameters were the daily heat loss, operating temperature, and the maximum discharge-charge capacity ratio. The optimum design for the storage unit in both regions had an operating temperature of 2400°C and 3% daily losses. Interestingly, there was a significant difference in the optimum discharge-charge ratio for the modeled regions. In the New England electric grid, the optimum storage unit had a maximum charging capacity of 1.5 times higher than the maximum discharging capacity. It was significantly higher in the ERCOT electric grid, where the maximum charging capacity was 3.75 times the maximum discharging capacity (333 % higher charging capacity). The reason is that the resulting cost-optimized grid in Texas has a higher share of solar PV installed, while the resulting grid in Massachusetts has a higher share of wind power. A solar-heavy grid requires storage units that can charge a large amount of energy during mid-day when there is a large amount of excess electricity from the solar PV plants and thereafter discharge over a longer period when there is a lack of solar availability. On the other hand, a wind-heavy electricity system does not require similar storage operation as wind power generates electricity more frequently through the day and does not have the typical mid-day generation pattern as solar energy. The significant difference in the optimum discharge-charge ratio indicates the importance of designing storage systems to match the specific conditions in each electricity market. In addition, we repeated the optimization schedule by incorporating specific details on the variation of the effective charge and discharge capacities as a function of SOC and the possibility

of obtaining negative SOC. The findings show that the engineering design was similar to the first design, but the cost improvement of TEGS compared to the baseline scenario was significantly lower. This motivates the research of developing solutions to maintain constant charging and discharging capacities as a function of SOC and avoid periods with negative SOC.

Once the optimum storage design was found, we investigated how TEGS affected the portfolio of supply technologies under different decarbonization scenarios. The results show that TEGS replaces Li-ion batteries and allows for a higher share of renewables and a lower share of fossil-fuel technologies in both the New England and ERCOT grid. In addition, to further assess the benefit of using TEGS to obtain cost-efficient decarbonization of the electric grid, we computed the shadow prices for different TEGS unit sizes as a function of storage duration. We find that TEGS positively improves the cost reduction of the electric power systems for all storage durations. In the end, we investigated how TEGS units can improve reliability in electric grids that are heavily decarbonized by modeling 22 different weather years. We found that TEGS vastly reduced the NSE cost compared to the baseline scenario for all weather years because of the increased amount of available power the TEGS unit can discharge to meet the demand when there is a lack of solar and wind availability. We show that cost-effective storage units have a critical role in maintaining resiliency in heavily decarbonized grids, which is critical to ensure a successful transition of the electric power system toward full decarbonization.

The findings show that design-optimized TES units can be essential in obtaining cost-efficient decarbonization of specific electric power systems while maintaining resiliency.

3 Experimental procedures

3.1 Electric power system modeling case studies

The case studies in this work cover two states in the U.S.: the New England grid in the state of Massachusetts and the Electric Reliability Council of Texas (ERCOT) grid in the state of Texas. We modeled the different regions as a "greenfield" system, that is, everything was built from scratch. In addition, we modeled each state as an idealized single-node system and did not include potential transmission constraints between the regions within each grid area. We are interested in modeling the potential of achieving fully decarbonized power systems, so we explore the year 2050. The electricity demand, capital cost, and performance data for the different generation technologies in these regions were collected from EIA and the National Renewable Energy Laboratory (NREL) using the annual technology baseline (ATB) with moderate cost assumptions for each technology [27]. The Github library PowerGenome [**] was used to collect the input data and shape them to the required format for the CEM. The weather year for modeling the VRE availability was 2021.

The reason for modeling the two states is to model the potential for using storage in different settings with different climates. More specifically, the state of Massachusetts is well known to have a large potential for wind power but less solar power owing to its high latitude. Conversely, Texas has a significantly larger solar resource, leading to a significantly different configuration for the cost-optimized future electric grid.

3.2 Optimizing the engineering design of TEGS

In this work, we are interested in investigating how the engineering design of TEGS can be optimized to enable full decarbonization of the electric power system cost-effectively. From the discussions in Section 1.2.1 in the Supplementary Material, two design parameters are of main importance for the optimization: daily heat loss and operating temperature. Here, we model a matrix in which the TEGS system has a daily heat loss between 1% and 10%, and a working temperature between 1900°C and 2400°C.

We want to find a TEGS configuration with the lowest cost when operating in a grid. TEGS costs are divided into Cost Per Power (CPP) and Cost Per Energy (CPE). CPP ($/kW) shows the capital cost for charging and discharging, while the CPE($/kWh) is the capital cost of the stored energy. The estimated capital cost of the TEGS unit is obtained from [15]. However, since TEGS is a technology currently under development, there are large uncertainties regarding the exact cost when fully integrated into the electric power system, and the capital cost is believed to be updated during the R&D process of the TEGS unit.

Fig. 6 shows how the cost of the TEGS system varies as a function of the temperature and heat loss.

[**]The PowerGenome GitHub library collects source data from the EIA, NREL, and EPA, and formats the input files for the CEM model. The GitHub library with associated documentation is found here `https://github.com/PowerGenome/PowerGenome`

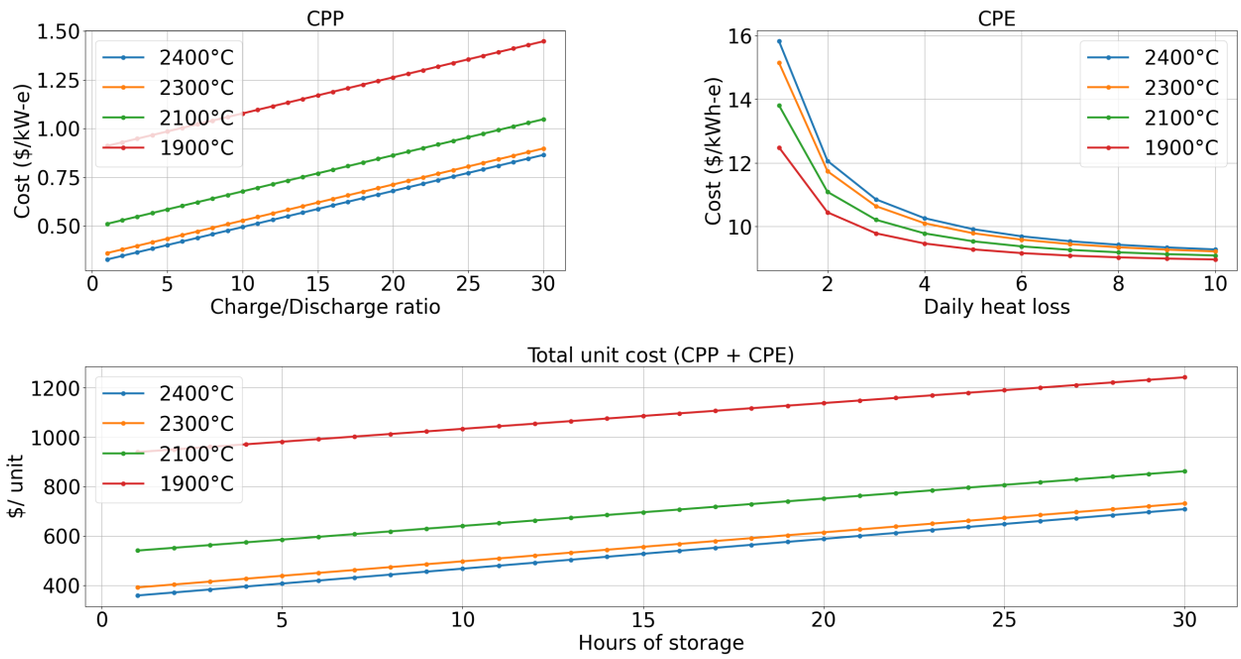


Figure 6: Different cost parameters of TEGS

Notably, the lowest temperature system (1900°C) has a lower CPE than the other designs. This is because of the lower need for system insulation when operating at lower temperatures. However, the CPP is significantly higher when the temperature is lower. This is because of the lower power density (higher energy density at higher temperatures). Therefore, more TPV equipment is required to discharge sufficient power to the electric power system. The lowermost figure combines the CPE and CPP costs, and the TEGS system is the cheapest at the highest possible temperatures.

From the description of the TEGS system in the Supplementary material, the power capacity of the TEGS system depends on the temperature during charging and discharging. During charging, the temperature increases (i.e., from 1900°C to 2400°C). This temperature increase increases the power density during charging and thus affects the power input during charging. Similarly, during discharging, the temperature decreases (i.e., from 2400°C to 1900°C), consequently decreasing the power density in the power block and affecting the power output from the TEGS unit during the discharge process. On the contrary, during charging, the power density increases as the temperature increases. Fig. 7 illustrate how the effective charge and discharge capacities change as a function of SOC.

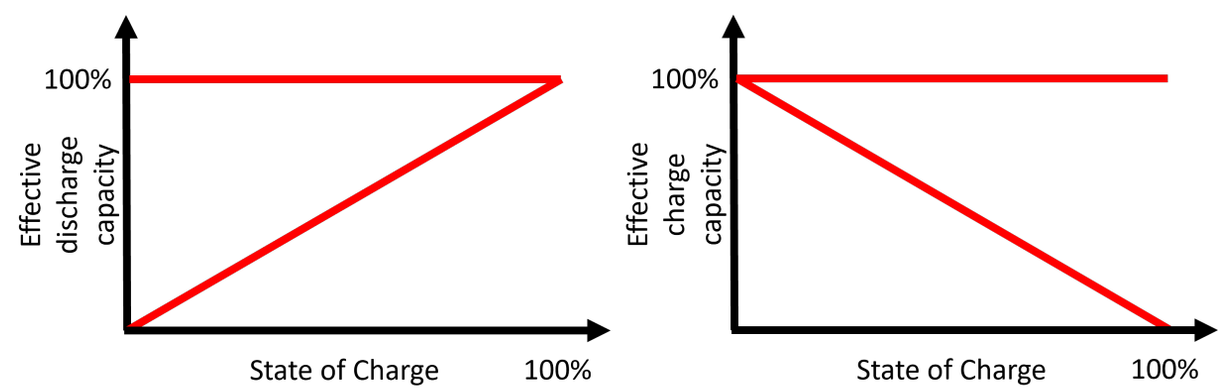


Figure 7: Charge and discharge profiles of TEGS as a function of SOC. The results obtained in Fig. 1 assume a constant charge and discharge profile, while the results obtained in Fig. 2 assume a charge and discharge profile that changes linearly as a function of SOC

Therefore, to model the importance of the relationship between the effective charge and discharge power and the SOC, we perform another modeling scenario where the effective charging and discharging capacities changes as a function of SOC. This represents that the temperature across the emitted TPV increases during charging, which increases the effective charging capacity. When discharging, the temperature across the emitted TPV is lower when the TEGS is depleted, so the effective discharge capacity decreases. We add the following constraints in the CEM to consider the changes in the effective charge and discharge capacities. The effective discharging capacity is defined as

$$\frac{\text{Effective discharging power}}{\text{Discharge capacity}} \leq \frac{\text{Stored energy}}{\text{Storage capacity}} \tag{1}$$

We implement a linear constraint in the CEM by transforming this equation into the following:

$$\frac{\text{Effective discharging power}}{\text{Discharge capacity}} \leq \frac{\text{Stored energy}}{\text{Discharge capacity * } c_1} \tag{2}$$

$$\rightarrow \text{Effective discharge power} * c_1 \leq \text{Stored energy} \tag{3}$$

where c_1 is a constant representing the discharge-storage ratio of the TEGS unit at different temperatures and heat losses. The discharge-storage ratio for each temperature and heat loss is obtained from the optimal ratio resulting from the modeling when assuming constant charging and discharging capacities (i.e., results in Fig. 1). Similarly the constraint for the effective charging capacity is implemented in the CEM as:

$$\frac{\text{Effective charging power}}{\text{Charge capacity}} \leq 1 - \frac{\text{Stored energy}}{\text{Storage capacity}} \tag{4}$$

$$\frac{\text{Effective charging power}}{\text{Charge capacity}} \leq 1 - \frac{\text{Stored energy}}{\text{Charging capacity * } c_2} \tag{5}$$

Where c_2 represents the charge-storage ratio of the optimum TEGS design with constant charge and discharge capacities. We use this equation to compute the effective charging capacity as a linear constraint:

$$\text{Effective charging capacity} * c_2 \leq \text{Charging capacity * } c_2 - \text{Stored energy} \tag{6}$$

In addition to incorporating effective charging and discharging capacities into the CEM, we add a third feature of the TEGS unit. The TEGS units have extremely high temperatures even when SOC = 0 (1900°C for a system of 2400°C when fully charged). Therefore, the TEGS unit will experience heat losses when SOC = 0, and consequently, the TEGS unit can obtain negative SOC if it is not recharged directly after being fully discharged.

We implement the effective discharge and charge capacities and the possibilities to obtain negative SOC to analyze these features' potential impact on the overall cost reduction. This will gain insight into whether these features are critical to tackle engineering-wise in order to obtain the required cost reduction and, thus, more cost-effectively enable full decarbonization of the electric power system.

3.3 Capacity Expansion Model (CEM)

The analysis utilizes GenX [26], an electric power system CEM that evaluates the cost-optimal electricity mix of generation, storage, and transmission infrastructure. Cost optimization is subject to several constraints, such as operational (electricity demand and generation) and policy (CO_2 emissions) constraints. The dominant constraints used in this study were the different electricity generation technologies available and the maximum limit on the allowed CO_2 emissions in the grid. We modeled different scenarios in which the following technologies were available: VRE generation (onshore and offshore wind power, and utility-scale solar PV), high-emission dispatchable power (Natural Gas combustion cycle, and Natural Gas combustion turbine), and low-emission dispatchable power (NG + CCS, and advanced fission). In addition, Li-ion and TEGS batteries can be used in different scenarios.

Here, the power system was modeled with a scenario without any CO_2 constraints (i.e., the model finds the cost-optimized electricity mix regardless of CO_2 emissions) and with scenarios in which the CO_2 emissions are reduced by 80%, 90%, 95%, and 99%. Constraining the maximum allowed CO_2 emissions in the grid will change the dynamics of how the power system is operated, as the grid can no longer use fossil-fuel-based technologies at the same scale. For each CO_2 reduction scenario, we modeled a baseline case in which TEGS was excluded from the model, and only Li ions were available as a storage facility in the grid. This allows for a comparison of how the cost of the grid changes when different configurations of the TEGS are modeled.

In total, when including modeling the TEGS design configurations with heat loss between 1% and 10% and operating temperature between 1900°C and 2400°C, 448 model runs are performed to obtain the results in this study.

To fully capture high-resolution temporal dependencies in the grid, we modeled the grid operation for each hour of the year. All scenarios were evaluated with the Gurobi optimization solver [28] using 16 cores with 128 GB RAM. All model scenarios were terminated with a 1% or lower optimality gap.

4 Acknowledgement

O.F.E. and M.C. acknowledge the support from the research project "Transformation to a Renewable & Smart Rural Power System Community (RENEW)", connected to the Arctic Centre for Sustainable Energy (ARC) at UiT-the Arctic University of Norway through Grant No. 310026.

References

[1] D. E. Remi Eriksen, Ulrike Haugen, Trond Hodne, Liv Hovem, Jin James Huang, Bent Erik Bakken, Kaveh Dianati, Thomas Horschig, Anne Louise Koefoed, Erica McConnell, Mats, S. S. Rinaldo, and Adrien Zambon. Energy transition outlook 2022, a global and regional forecast to 2050, 2022.

[2] Dolf Gielen, Francisco Boshell, Deger Saygin, Morgan D Bazilian, Nicholas Wagner, and Ricardo Gorini. The role of renewable energy in the global energy transformation. *Energy Strategy Reviews*, 24:38–50, 2019.

[3] Harry Apostoleris, Sgouris Sgouridis, Marco Stefancich, and Matteo Chiesa. Utility solar prices will continue to drop all over the world even without subsidies. *Nature Energy*, 4(10):833–834, 2019.

[4] William A Braff, Joshua M Mueller, and Jessika E Trancik. Value of storage technologies for wind and solar energy. *Nature Climate Change*, 6(10):964–969, 2016.

[5] Micah S Ziegler, Joshua M Mueller, Gonçalo D Pereira, Juhyun Song, Marco Ferrara, Yet-Ming Chiang, and Jessika E Trancik. Storage requirements and costs of shaping renewable energy toward grid decarbonization. *Joule*, 3(9):2134–2153, 2019.

[6] Matthew R Shaner, Steven J Davis, Nathan S Lewis, and Ken Caldeira. Geophysical constraints on the reliability of solar and wind power in the united states. *Energy & Environmental Science*, 11(4):914–925, 2018.

[7] Nestor A Sepulveda, Jesse D Jenkins, Aurora Edington, Dharik S Mallapragada, and Richard K Lester. The design space for long-duration energy storage in decarbonized power systems. *Nature Energy*, 6(5):506–516, 2021.

[8] Xinhai Xu, K Vignarooban, Ben Xu, Keng Hsu, and Arunachala Mada Kannan. Prospects and problems of concentrating solar power technologies for power generation in the desert regions. *Renewable and Sustainable Energy Reviews*, 53:1106–1131, 2016.

[9] Kathleen M Kennedy, Tyler H Ruggles, Katherine Rinaldi, Jacqueline A Dowling, Lei Duan, Ken Caldeira, and Nathan S Lewis. The role of concentrated solar power with thermal energy storage in least-cost highly reliable electricity systems fully powered by variable renewable energy. *Advances in Applied Energy*, 6:100091, 2022.

[10] Micah S Ziegler and Jessika E Trancik. Re-examining rates of lithium-ion battery technology improvement and cost decline. *Energy & Environmental Science*, 14(4):1635–1651, 2021.

[11] Dharik S Mallapragada, Nestor A Sepulveda, and Jesse D Jenkins. Long-run system value of battery energy storage in future grids with increasing wind and solar generation. *Applied Energy*, 275:115390, 2020.

[12] Robert Armstrong, Yet-Ming Chiang, Howard Gruenspecht, Fikile Brushett, John Deutch, Seiji Engelkemier, Emgre Gencer, Robert Jaffe, Paul Joskow, Dharik Mallapragada, Elsa Olivetti, Richard Schmalensee, Robert Stoner, Chi-Jen Yang, Bjorn Brandtzaeg, Patrick Brown, Kevin Huang, Johannes Pfeifenberger, Francis O'Sullivan, and Yang Shao-Horn. *The Future of Energy storage*. MIT Future of series. Massachusetts Institute of Technology, 2022.

[13] Paul Albertus, Joseph S Manser, and Scott Litzelman. Long-duration electricity storage applications, economics, and technologies. *Joule*, 4(1):21–32, 2020.

[14] Caleb Amy, Hamid Reza Seyf, Myles A Steiner, Daniel J Friedman, and Asegun Henry. Thermal energy grid storage using multi-junction photovoltaics. *Energy & Environmental Science*, 12(1):334–343, 2019.

[15] Colin C Kelsall, Kyle Buznitsky, and Asegun Henry. Technoeconomic analysis of thermal energy grid storage using graphite and tin. *arXiv preprint arXiv:2106.07624*, 2021.

[16] Asegun Henry, Ravi Prasher, and Arun Majumdar. Five thermal energy grand challenges for decarbonization. *Nature Energy*, 5(9):635–637, 2020.

[17] Alina LaPotin, Kevin L Schulte, Myles A Steiner, Kyle Buznitsky, Colin C Kelsall, Daniel J Friedman, Eric J Tervo, Ryan M France, Michelle R Young, Andrew Rohskopf, et al. Thermophotovoltaic efficiency of 40%. *Nature*, 604(7905):287–291, 2022.

[18] Wolf-Peter Schill and Alexander Zerrahn. Long-run power storage requirements for high shares of renewables: Results and sensitivities. *Renewable and Sustainable Energy Reviews*, 83:156–171, 2018.

[19] Fernando J De Sisternes, Jesse D Jenkins, and Audun Botterud. The value of energy storage in decarbonizing the electricity sector. *Applied Energy*, 175:368–379, 2016.

[20] Clara F Heuberger, Iain Staffell, Nilay Shah, and Niall Mac Dowell. A systems approach to quantifying the value of power generation and energy storage technologies in future electricity networks. *Computers & Chemical Engineering*, 107:247–256, 2017.

[21] Nestor A Sepulveda, Jesse D Jenkins, Fernando J de Sisternes, and Richard K Lester. The role of firm low-carbon electricity resources in deep decarbonization of power generation. *Joule*, 2(11):2403–2420, 2018.

[22] Bethany A Frew, Sarah Becker, Michael J Dvorak, Gorm B Andresen, and Mark Z Jacobson. Flexibility mechanisms and pathways to a highly renewable us electricity future. *Energy*, 101:65–78, 2016.

[23] Hailiang Liu, Tom Brown, Gorm Bruun Andresen, David P Schlachtberger, and Martin Greiner. The role of hydro power, storage and transmission in the decarbonization of the chinese power system. *Applied Energy*, 239:1308–1321, 2019.

[24] Mehdi Jafari, Magnus Korpås, and Audun Botterud. Power system decarbonization: Impacts of energy storage duration and interannual renewables variability. *Renewable Energy*, 156:1171–1185, 2020.

[25] Anna H Schleifer, Caitlin A Murphy, Wesley J Cole, and Paul Denholm. Exploring the design space of pv-plus-battery system configurations under evolving grid conditions. *Applied Energy*, 308:118339, 2022.

[26] Jesse D Jenkins and Nestor A Sepulveda. Enhanced decision support for a changing electricity landscape: the genx configurable electricity resource capacity expansion model. 2017.

[27] NREL (National Renewable Energy Laboratory). 2022 annual technology baseline, 2022.

[28] Miles Lubin and Iain Dunning. Computing in operations research using julia. *INFORMS Journal on Computing*, 27(2):238–248, 2015.

COST-EFFECTIVE THERMAL ENERGY GRID STORAGE FOR DECARBONIZING ELECTRIC POWER SYSTEMS

Odin Foldvik Eikeland[1,3,*], Ruaridh Macdonald[2,*], Harry Apostoleris[4], Shomik Verma[1], Kyle Buznitsky[1], Matteo Chiesa[1,3,5,], & Asegun Henry[1,**]**

[1]Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA.
[2]MIT Energy Initiative, Massachusetts Institute of Technology, Cambridge, MA, USA.
[3]Department of Physics and Technology, UiT – the Arctic University of Norway, 9037 Tromsø, Norway.
[4]Dubai Electricity & Water Authority
[5]Laboratory for Energy and NanoScience (LENS), Khalifa University of Science and Technology, Masdar.

1 Supplementary material

1.1 Extended results

1.1.1 Optimized TEGS design

Fig. 1 shows the resulting optimized TEGS under different CO_2 scenarios for the ERCOT power system. The results are similar to the New England power system in the main manuscript. The optimum TEGS configuration that achieves the greatest cost reduction in the Texas region has an operating temperature of 2400°C and a daily heat loss of approximately 3%. Like the New England grid, the optimum configuration results in approximately 3% cost reduction compared to the baseline case where TEGS is excluded from the grid. In the third contour plot to the right, the charge-discharge ratio is approximately 3.75, showing that the TEGS unit should have a charging capacity that is almost 4 times higher than the discharging capacity. This is significantly higher in the New England grid, where the storage unit should charge with a maximum capacity that is 2 times higher than the maximum discharging capacity. The difference in charging requirements for the two different power systems is because the share of solar energy in the ERCOT grid is significantly higher than in the New England grid, where wind power dominates technology. In a solar-heavy power system, the periodic generation profile of the solar PV systems gives excess electricity during mid-day when the sun is shining but less electricity during the morning and afternoon when the sun is below the horizon. This requires storage systems to be able to charge at an extremely high capacity when there is excess electricity during mid-day. Then the stored energy can be discharged over a longer period when there is a lack of solar availability. Due to the periodicity of electricity generation from solar energy, approximately 30 hours of storage are needed in the ERCOT grid instead of 50 hours for the New England grid.

As in the main manuscript, we performed another modeling procedure considering the effective charge and discharge capacities and the possibility of obtaining a negative SOC. The results are presented in Fig. 2.

The results show a similar optimum design for the TEGS unit as in Fig. 1. However, considering the effective charge and discharge capacities and the possibility of obtaining a negative SOC significantly impact the value of using TEGS in the ERCOT grid. Now there is no longer a cost reduction using TEGS, and the baseline scenario where Li-ion batteries are the only storage unit available is the preferred solution with respect to cost-effectively decarbonize the ERCOT grid. Similar to the discussions in the main manuscript concerning the New England grid, the large impact on the cost reductions highlights the importance of researching solutions to maintain a constant capacity during charge and discharge and avoid periods with negative SOC.

[*]These authors contributed equally to this work
[**]Corresponding authors: mchiesa@mit.edu and ase@mit.edu

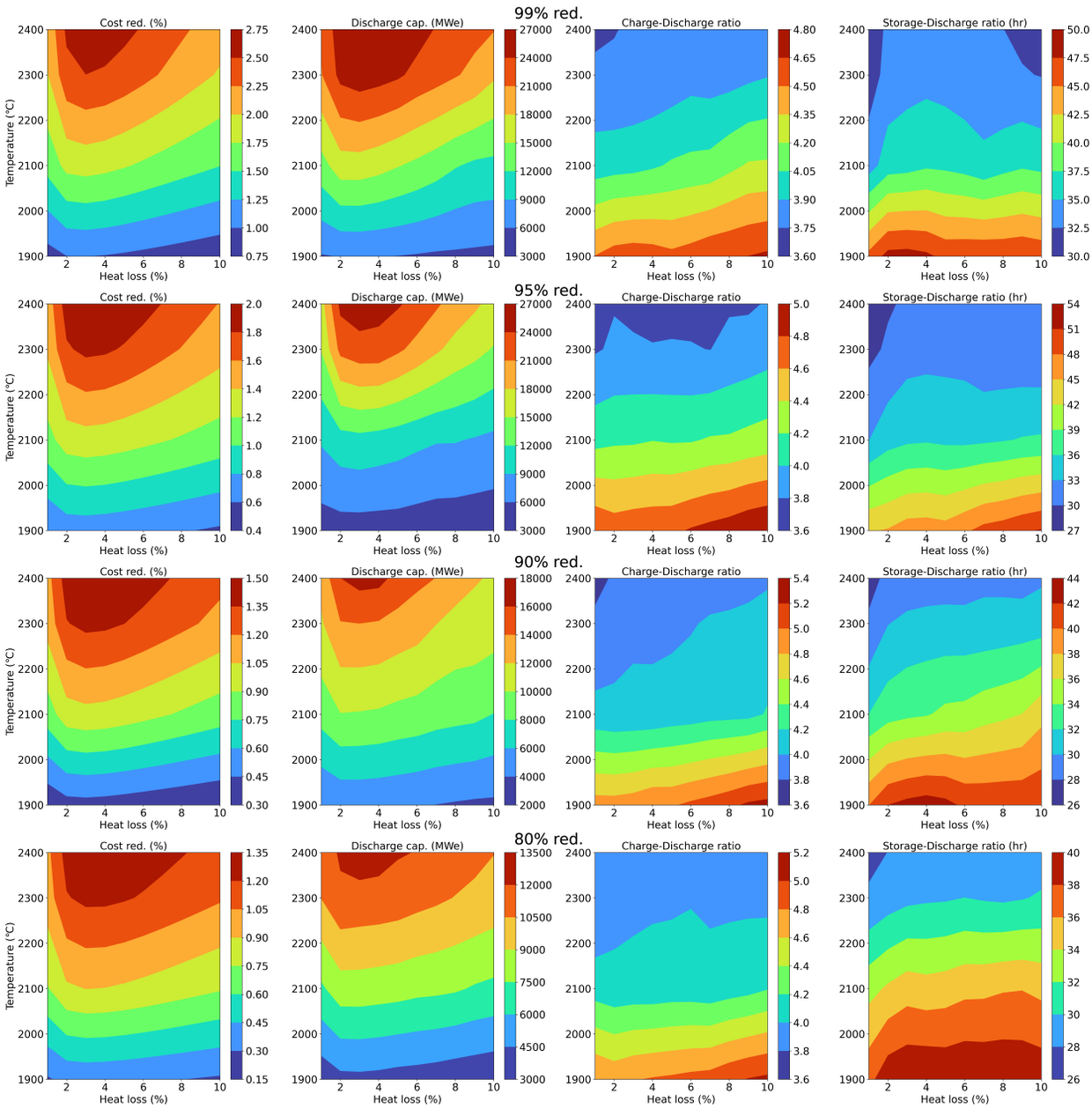


Figure 1: Optimized TEGS configurations under different CO_2 reduction scenarios in the ERCOT grid.

1.2 Shadow prices

As for the main manuscript, Fig. 3 shows the evolution of the shadow price for the optimum TEGS design at different discharge capacities.

The shadow prices show similar results to the New England grid, where at least 20 hours of storage is required to reduce the costs of the ERCOT grid using TEGS. The shadow prices of charging are indicated in the upper right graph, showing the same trend as in the main manuscript, as the shadow prices reduce for longer storage durations. The lower graph shows that adding TEGS units positively impacts the cost reductions of the ERCOT grid for all storage durations and all discharge capacities.

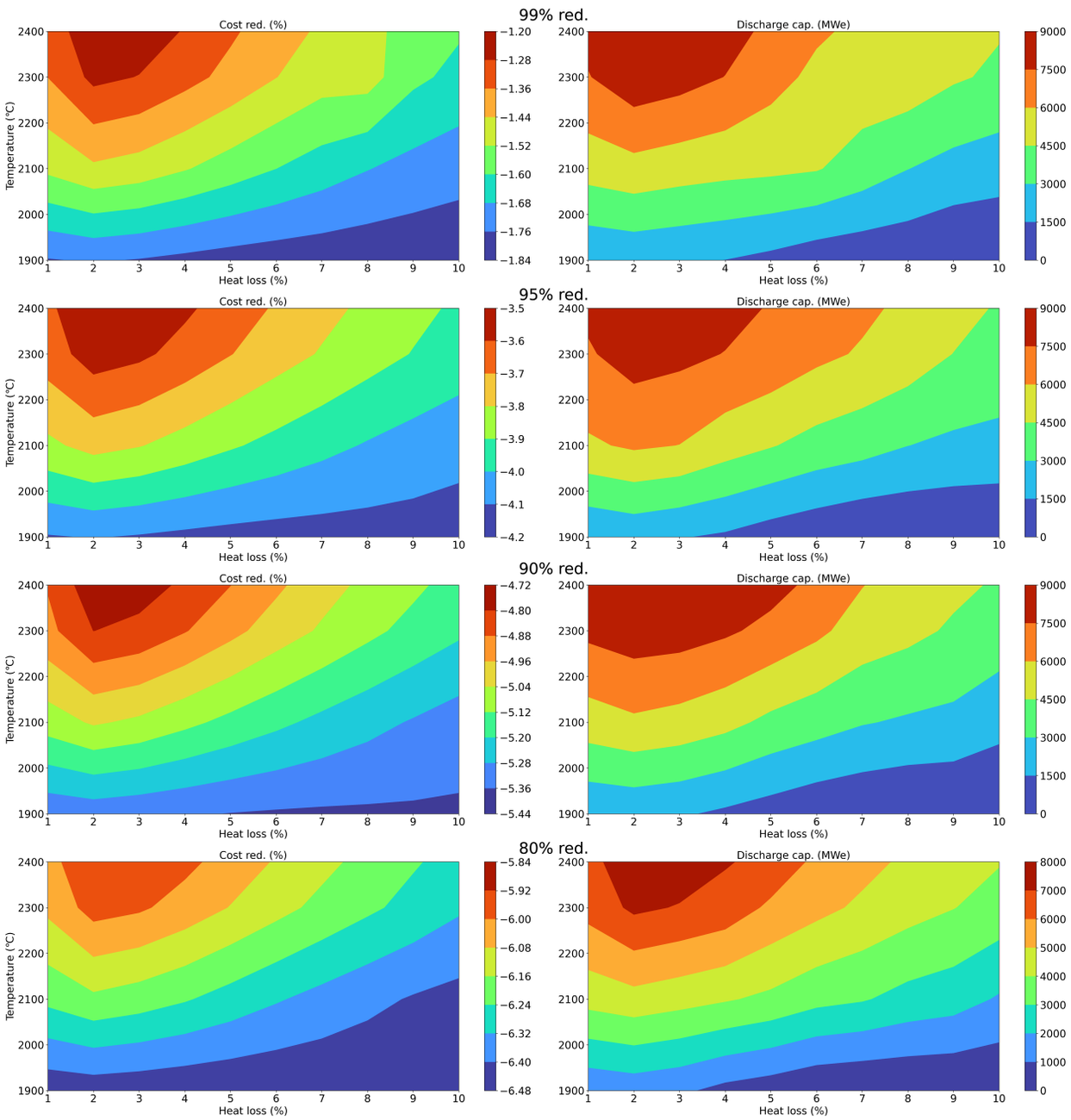


Figure 2: Optimized TEGS configurations under different CO_2 reduction scenarios in the in the ERCOT grid considering the effective discharge and charge capacities, and the possibilities of obtaining negative SOC.

1.2.1 Weekly operating examples

This section provides insight into the hourly operation of the New England and ERCOT electricity systems, respectively. This is useful to better understand how an electric power system operates hour-by-hour and gives insight into how storage is utilized to cover the electricity demand when the supply from VRE sources is insufficient. Fig. 4 shows the hourly operation of the modeled grid under the 99% CO_2 reduction scenario in the New England grid. The uppermost Figure shows a winter-week example, while the lowermost figure shows a summer-week example. Wind power and Natural Gas with CCS is the dominant source of electricity supply to meet demand.

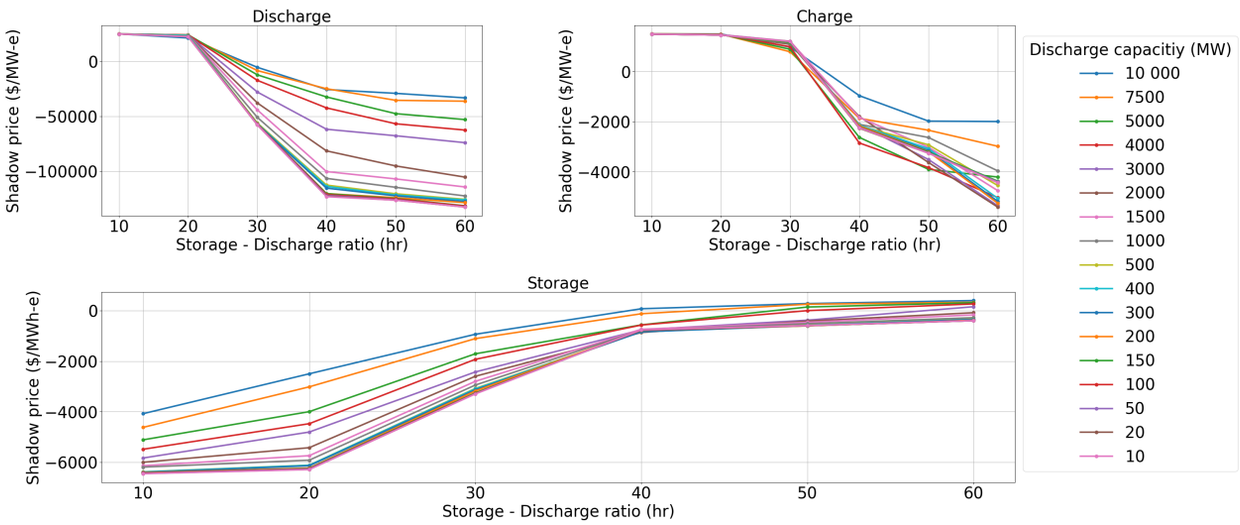


Figure 3: Shadow prices of TEGS in the ERCOT grid

In Fig. 4, the electricity demand must be covered is illustrated as a dotted blue line, whereas the dotted red line shows the net generation from renewables (solar + wind) and the discharged power from the Li-ion battery and TEGS. The stacked colors show how each technology contributes to covering the total electricity demand. Clearly, the dominant source of electricity generation is land-based wind power in addition to Natural Gas with CCS when there is lower wind availability. Interestingly, for the winter week example, the electricity demand is 100%, covered by wind power generation alone in the last two days. In these days, the excess electricity generation from wind power is used to charge the TEGS storage unit. As CO_2 emissions are heavily constrained, there are minor contributions from NG that have associated emissions. The NG technology that is actively providing power to the grid whenever there is insufficient wind power is NG with 100% CCS and 90% CCS (NG-CCCCS). Interestingly, when wind power is unavailable, TEGS significantly covers the demand by discharging large amounts of power. In the supplementary, weekly illustrations for the Texas grid are provided. Here, it is clear that solar energy is the dominating source of energy supply, and the storage unit plays a more active role in discharging power when there is a lack of solar availability.

The red dotted line, which shows the net supply from renewables and discharged power from storage, indicates that the grid receives electricity nearly 100% of the time. This shows the benefits of using storage to improve the dispatchability of renewables with an intermittent generation profile.

Fig. 5 shows the hourly operation of the modeled grid under the 99% CO_2 reduction scenario in the ERCOT grid. In contradiction to the New England region, where the whole demand could be covered with wind power for some periods during the winter, the electricity generation from solar PV plants plays a more significant role. However, when the sun sets, the technology that contributes the most to covering the overall electricity demand is the power discharged from TEGS. It is clear that NG and biomass play secondary roles in contributing to covering the total demand. A similar pattern occurs during the summer week example.

1.3 Experimental procedures

1.3.1 Thermal Energy Grid Storage (TEGS)

Fig. 6 illustrates the different components of the TEGS concept [1, 2].

Excess electricity is used to fuel resistive heating materials (graphite) to charge the TEGS unit, transforming the electricity into heat at a temperature exceeding 2500°C. Then, the energy is transferred to graphite conduits via thermal radiation. Inside the pipes, liquid tin is used as the heat transfer fluid. The tin is heated from 1900°C to 2400°C, transforming the energy input into sensible heat and increasing its enthalpy. The liquid tin is continuously pumped through the conduits and then conveyed to the graphite blocks in the storage unit. When the 2400°C tin is pumped through the graphite blocks via pipes, it heats them from 1900°C to 2400°C via thermal radiation. Consequently, this cools the tin back to 1900°C. The tin is then reheated by being pumped back through the resistance heaters. This process constitutes the charging process until the graphite blocks are heated back to peak temperature. The storage unit should have sufficiently large thermal mass to enable the storage unit to be charged for long periods with low

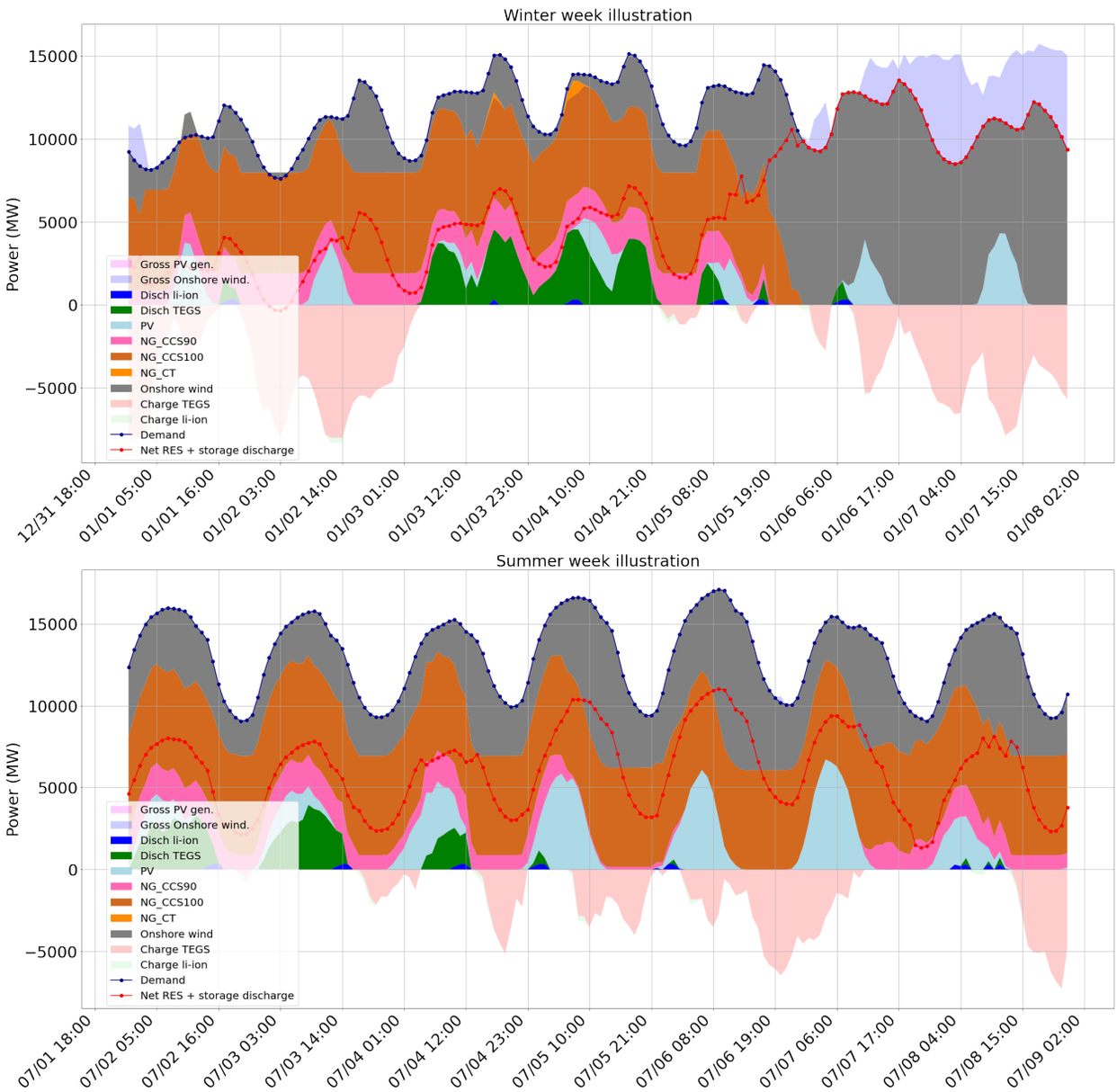


Figure 4: One week of hourly operation in the New England grid

heat loss (i.e., < 10%) [1]. Heat loss is a crucial design consideration in finding the most economically advantageous TEGS unit, and this study models a TEGS unit with heat loss between 1% and 10%. During discharging, liquid tin is pumped through the graphite storage to a power block. The power block consists of graphite conduits with unit cells. Each unit cell of piping creates a rectangular cavity lined with tungsten foil. This is a diffusion barrier to prevent graphite deposition onto the TPV cells. Inside each cavity, the TPV cells can be lowered into the unit cell cavity. Here the TPV cells will be illuminated with the light emitted by the tungsten foil, which is heated by the light emitted by the graphite conduit carrying the tin. This net transfer of energy converts a large fraction (> 50%) of the energy to electricity, which causes the tin's temperature to decrease to 1900°C before being pumped back to the graphite storage unit, where the tin is reheated again during the charging phase. The system is heated up to 2400°C to increase the power density in the power block during discharge to maximize the discharge capacity of the system. However, achieving such extreme temperatures is difficult engineering-wise, and investigating the possibility of reducing the required operating temperature is another crucial design choice modeled in this study.

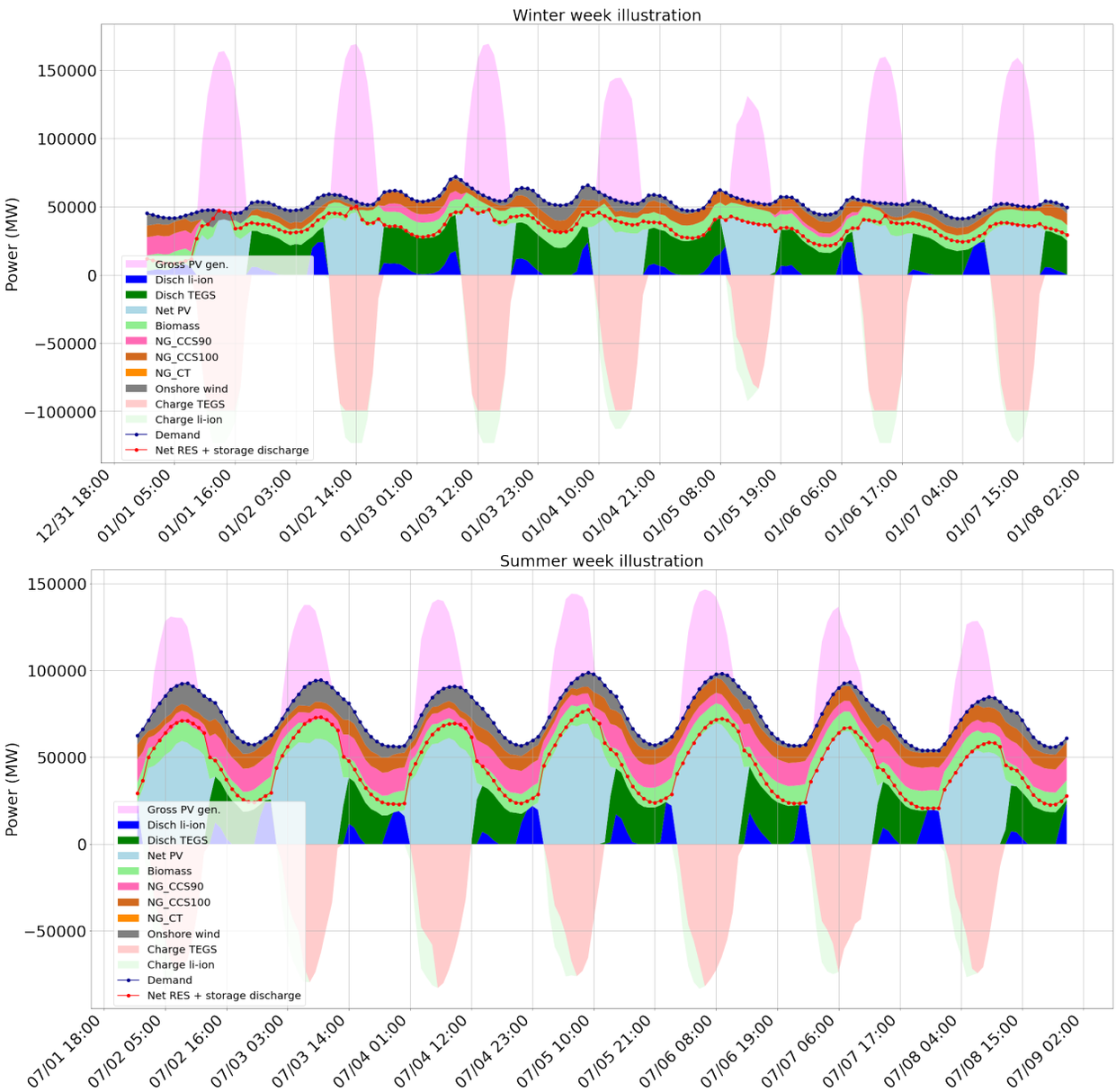


Figure 5: One week of hourly operation in the ERCOT grid

However, as the temperature in the power block decreases during discharge (i.e., from 2400°C to 1900°C), the power density consequently decreases. This can affect the effective discharge capacity from the TEGS unit as the power density in the power block decreases during the discharge process. Similarly, as the temperature in the power block during charging increases, the effective charging capacity of the TEGS unit increases during the charging process. The variation of the effective charge and discharge capacities as a function of the State of Charge is an additional design parameter that is modeled in this study. The TPV conversion efficiency (i.e., the discharge efficiency from heat to electricity) determines the TEGS unit's RTE. Recent discoveries in [3] enabled a TPV efficiency of 50%, giving an RTE of 50%. The charging efficiency (from heat to electricity) is assumed to be 100%. In summary, the TEGS unit is a rechargeable grid-scale thermal battery that can store energy as heat and supply electricity to the grid on demand with a projected cost sufficiently low to enable full decarbonization of the electric power system.

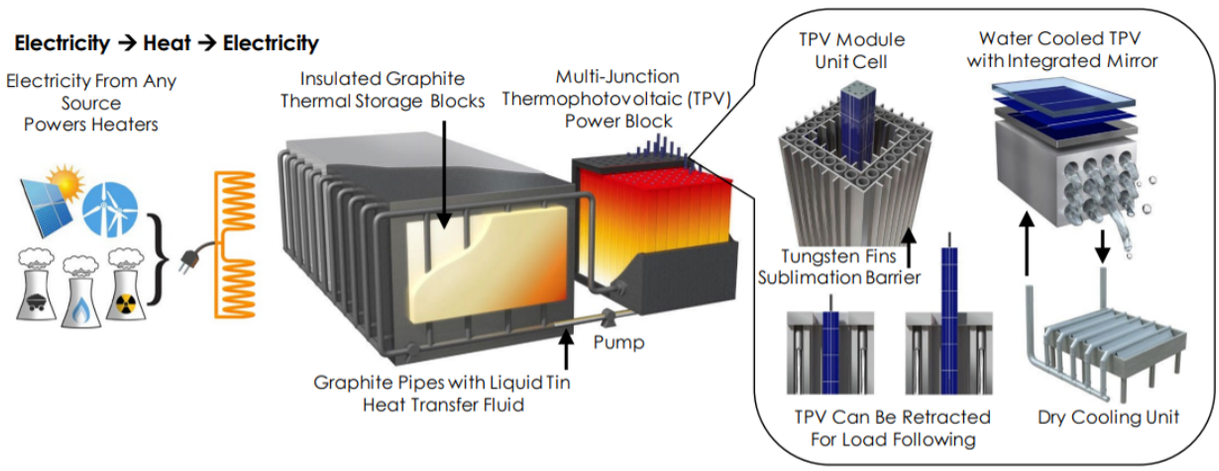


Figure 6: Schematic illustration of TEGS components. Illustration collected from [1]

References

[1] Colin C Kelsall, Kyle Buznitsky, and Asegun Henry. Technoeconomic analysis of thermal energy grid storage using graphite and tin. *arXiv preprint arXiv:2106.07624*, 2021.

[2] Caleb Amy, Hamid Reza Seyf, Myles A Steiner, Daniel J Friedman, and Asegun Henry. Thermal energy grid storage using multi-junction photovoltaics. *Energy & Environmental Science*, 12(1):334–343, 2019.

[3] Alina LaPotin, Kevin L Schulte, Myles A Steiner, Kyle Buznitsky, Colin C Kelsall, Daniel J Friedman, Eric J Tervo, Ryan M France, Michelle R Young, Andrew Rohskopf, et al. Thermophotovoltaic efficiency of 40%. *Nature*, 604(7905):287–291, 2022.

References

Ahmad, T., & Chen, H. (2020). A review on machine learning forecasting growth trends and their real-time applications in different energy systems. *Sustainable Cities and Society*, *54*, 102010.

Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017). Understanding of a convolutional neural network. In *2017 international conference on engineering and technology (icet)* (pp. 1–6).

Albertus, P., Manser, J. S., & Litzelman, S. (2020). Long-duration electricity storage applications, economics, and technologies. *Joule*, *4*(1), 21–32.

Almas, M. S., Vanfretti, L., Løvlund, S., & Gjerde, J. (2014). Open source scada implementation and pmu integration for power system monitoring and control applications. In *2014 ieee pes general meeting| conference & exposition* (pp. 1–5).

Alpaydin, E. (2020). *Introduction to machine learning*. MIT press.

Amy, C., Seyf, H. R., Steiner, M. A., Friedman, D. J., & Henry, A. (2019). Thermal energy grid storage using multi-junction photovoltaics. *Energy & Environmental Science*, *12*(1), 334–343.

Apostoleris, H., Sgouridis, S., Stefancich, M., & Chiesa, M. (2018). Evaluating the factors that led to low-priced solar electricity projects in the middle east. *Nature Energy*, *3*(12), 1109–1114.

Armstrong, R., Chiang, Y.-M., Gruenspecht, H., Brushett, F., Deutch, J., Engelkemier, S., ... Shao-Horn, Y. (2022). *The future of energy storage*. Massachusetts Institute of Technology.

Athmaja, S., Hanumanthappa, M., & Kavitha, V. (2017). A survey of machine learning algorithms for big data analytics. In *2017 international conference on innovations in information, embedded and communication systems (iciiecs)* (pp. 1–4).

Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.-R., & Samek, W. (2015). On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PloS one*, *10*(7), e0130140.

Baik, E., Chawla, K. P., Jenkins, J. D., Kolster, C., Patankar, N. S., Olson, A., ... Long, J. C. (2021). What is different about different net-zero carbon electricity systems? *Energy and Climate Change*, *2*, 100046.

Bebis, G., & Georgiopoulos, M. (1994). Feed-forward neural networks. *Ieee Potentials*, *13*(4), 27–31.

Bedi, J., & Toshniwal, D. (2019). Deep learning framework to forecast electricity demand. *Applied energy*, *238*, 1312–1326.

Belagoune, S., Bali, N., Bakdi, A., Baadji, B., & Atif, K. (2021). Deep learning through lstm classification and regression for transmission line fault detection, diagnosis and location in large-scale multi-machine power systems. *Measurement*, *177*, 109330.

Belyadi, H., & Haghighat, A. (2021). *Machine learning guide for oil and gas using python: A step-by-step breakdown with data, algorithms, codes, and*

applications. Gulf Professional Publishing.

Bhattarai, B. P., Paudyal, S., Luo, Y., Mohanpurkar, M., Cheung, K., Tonkoski, R., ... others (2019). Big data analytics in smart grids: state-of-the-art, challenges, opportunities, and future directions. *IET Smart Grid*, *2*(2), 141–154.

Bianchi, F. M., Maiorino, E., Kampffmeyer, M. C., Rizzi, A., & Jenssen, R. (2017). An overview and comparative analysis of recurrent neural networks for short term load forecasting. *arXiv preprint arXiv:1705.04378*.

Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4) (No. 4). Springer.

Bloomberg. (2023). *How artificial intelligence is fueling the energy transition*. Retrieved from `https://sponsored.bloomberg.com/article/BHC3/how-artificial-intelligence-is-fueling-the-energy-transition`

Blume, S. W. (2016). *Electric power system basics for the nonelectrical professional*. John Wiley & Sons.

Bocca, R., & Ashraf, M. (2022). *Net-zero industry tracker*.

Boroojeni, K. G., Amini, M. H., & Iyengar, S. S. (2017). *Smart grids: security and privacy issues* (Vol. 221). Springer.

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on computational learning theory* (pp. 144–152).

Bouckaert, S., Pales, A. F., McGlade, C., Remme, U., Wanner, B., Varro, L., ... Spencer, T. (2021). Net zero by 2050: A roadmap for the global energy sector.

Bouktif, S., Fiaz, A., Ouni, A., & Serhani, M. A. (2018). Optimal deep learning lstm model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches. *Energies*, *11*(7), 1636.

Braff, W. A., Mueller, J. M., & Trancik, J. E. (2016). Value of storage technologies for wind and solar energy. *Nature Climate Change*, *6*(10), 964–969.

Breyer, C., Khalili, S., Bogdanov, D., Ram, M., Oyewo, A. S., Aghahosseini, A., ... others (2022). On the history and future of 100% renewable energy systems research. *IEEE Access*, *10*, 78176–78218.

Chen, K., Hu, J., Zhang, Y., Yu, Z., & He, J. (2019). Fault location in power distribution systems via deep graph convolutional networks. *IEEE Journal on Selected Areas in Communications*, *38*(1), 119–131.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

Chou, J.-S., & Tran, D.-S. (2018). Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders. *Energy*, *165*, 709–726.

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.

Clifford, C. (2023). *Why america's outdated energy grid is a climate problem*. Retrieved 2023-03-08, from https://www.cnbc.com/2023/02/17/why-americas-outdated-energy-grid-is-a-climate-problem.html

Cole, W. J., Greer, D., Denholm, P., Frazier, A. W., Machen, S., Mai, T., ... Baldwin, S. F. (2021). Quantifying the challenge of reaching a 100% renewable energy power system for the united states. *Joule*, *5*(7), 1732–1748.

De Benedetti, M., Leonardi, F., Messina, F., Santoro, C., & Vasilakos, A. (2018). Anomaly detection and predictive maintenance for photovoltaic systems. *Neurocomputing*, *310*, 59–68.

Denholm, P., Arent, D. J., Baldwin, S. F., Bilello, D. E., Brinkman, G. L., Cochran, J. M., ... others (2021). The challenges of achieving a 100% renewable electricity system in the united states. *Joule*, *5*(6), 1331–1352.

Dimmen, S. D. (2022). *Kunstig intelligens skal forutsi strømbehovet*. Retrieved from https://www.nrk.no/tromsogfinnmark/gjor-stromnettet-pa-senja-smartere-med-kunstig-intelligens-_-sparer-naeringslivet-for-milliontap-1.15810933

Dixon, M. F., Halperin, I., & Bilokon, P. (2020). *Machine learning in finance* (Vol. 1170). Springer.

Donti, P. L., & Kolter, J. Z. (2021). Machine learning for sustainable energy systems. *Annual Review of Environment and Resources*, *46*, 719–747.

Duchesne, L., Karangelos, E., & Wehenkel, L. (2020). Recent developments in machine learning for energy systems reliability management. *Proceedings of the IEEE*, *108*(9), 1656–1676.

D'Aprile, A. (2018). Arctic tourism: How the great north is becoming the new exotic. *Future Earth*.

Eikeland, O. F., Apostoleris, H., Santos, S., Ingebrigtsen, K., Boström, T., & Chiesa, M. (2020). Rethinking the role of solar energy under location specific constraints. *Energy*, *211*, 118838.

Eikeland, O. F., Bianchi, F. M., Holmstrand, I. S., Bakkejord, S., & Chiesa, M. (2022). Detecting the linear and non-linear causal links for disturbances in the power grid. In *Intelligent technologies and applications: 4th international conference, intap 2021, grimstad, norway, october 11–13, 2021, revised selected papers* (pp. 325–336).

Eikeland, O. F., Holmstrand, I. S., Bakkejord, S., Chiesa, M., & Bianchi, F. M. (2021). Detecting and interpreting faults in vulnerable power grids with machine learning. *IEEE Access*, *9*, 150686–150699.

Eikeland, O. F., Hovem, F. D., Olsen, T. E., Chiesa, M., & Bianchi, F. M. (2022). Probabilistic forecasts of wind power generation in regions with complex topography using deep learning methods: An arctic case. *Energy Conversion and Management: X*, 100239.

Elman, J. L. (1990). Finding structure in time. *Cognitive science*, *14*(2), 179–211.

Enel. (2023). *3 ways ai is accelerating the energy transition.* Retrieved from https://openinnovability.enel.com/media/insights/2023/02/how-ai-accelerates-energy-transition

EPA. (2011). *Inventory of us greenhouse gas emissions and sinks: 1990–2009,*.

Eriksen, D. E. R., Haugen, U., Hodne, T., Hovem, L., Huang, J. J., Bakken, B. E., ... Zambon, A. (2022). *Energy transition outlook 2022, a global and regional forecast to 2050*.

EU. (2020). *A european green deal.* Retrieved from https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en

Fazlollahtabar, H., Saidi-Mehrabad, M., & Balakrishnan, J. (2015). Mathematical optimization for earliness/tardiness minimization in a multiple automated guided vehicle manufacturing system via integrated heuristic algorithms. *Robotics and Autonomous Systems*, *72*, 131–138.

Foldvik Eikeland, O., Bianchi, F. M., Apostoleris, H., Hansen, M., Chiou, Y.-C., & Chiesa, M. (2021a). Predicting energy demand in semi-remote arctic locations. *Energies*, *14*(4). Retrieved from https://www.mdpi.com/1996-1073/14/4/798 doi: 10.3390/en14040798

Foldvik Eikeland, O., Bianchi, F. M., Apostoleris, H., Hansen, M., Chiou, Y.-C., & Chiesa, M. (2021b). Predicting energy demand in semi-remote arctic locations. *Energies*, *14*(4), 798.

Foley, A. M., Leahy, P. G., Marvuglia, A., & McKeogh, E. J. (2012). Current methods and advances in forecasting of wind power generation. *Renewable energy*, *37*(1), 1–8.

Gasparin, A., Lukovic, S., & Alippi, C. (2022). Deep learning for time series forecasting: The electric load case. *CAAI Transactions on Intelligence Technology*, *7*(1), 1–25.

Gazzea, M., Aalhus, S., Kristensen, L. M., Ozguven, E. E., & Arghandeh, R. (2021). Automated 3d vegetation detection along power lines using monocular satellite imagery and deep learning. In *2021 ieee international geoscience and remote sensing symposium igarss* (pp. 3721–3724).

Golub, G. H., Hansen, P. C., & O'Leary, D. P. (1999). Tikhonov regularization and total least squares. *SIAM journal on matrix analysis and applications*, *21*(1), 185–194.

Graves, A. (2011). Practical variational inference for neural networks. *Advances in neural information processing systems*, *24*.

Graves, A. (2013). Generating sequences with recurrent neural networks. *arXiv preprint arXiv:1308.0850*.

Greer, M. (2012). Chapter 3 - u.s. electric markets, structure, and regulations. In M. Greer (Ed.), *Electricity marginal cost pricing* (p. 39-100). Boston: Butterworth-Heinemann. Retrieved from https://www.sciencedirect.com/science/article/pii/B978012385134500003X doi: https://doi.org/10.1016/B978-0-12-385134-5.00003-X

Grigsby, L. L. (2007). *Electric power generation, transmission, and distribution*. CRC press.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., ... others (2018). Recent advances in convolutional neural networks. *Pattern recognition*, *77*, 354–377.

Harrell, F. E., et al. (2001). *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis* (Vol. 608). Springer.

Henry, A., Prasher, R., & Majumdar, A. (2020). Five thermal energy grand challenges for decarbonization. *Nature Energy*, *5*(9), 635–637.

Hezam, I. M., & Nayeem, M. K. (2020). A systematic literature review on mathematical models of humanitarian logistics. *Symmetry*, *13*(1), 11.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735–1780.

Hoffmann, V., Michałowska, K., Andresen, C., & Torsæter, B. N. (2019). Incipient fault prediction in power quality monitoring.

Hong, T., & Fan, S. (2016). Probabilistic electric load forecasting: A tutorial review. *International Journal of Forecasting*, *32*(3), 914–938.

Hossain, M. A., Chakrabortty, R. K., Elsawah, S., Gray, E. M., & Ryan, M. J. (2021). Predicting wind power generation using hybrid deep learning with optimization. *IEEE Transactions on Applied Superconductivity*, *31*(8), 1–5.

Hu, Y., Huber, A., Anumula, J., & Liu, S.-C. (2018). Overcoming the vanishing gradient problem in plain recurrent networks. *arXiv preprint arXiv:1801.06105*.

IPCC. (2014). *Ipcc climate change 2014: Synthesis report.*

IPCC. (2018). *Sr15 – special report: Global warming of 1.5c.* Retrieved 2023-03-08, from `https://www.ipcc.ch/sr15/`

IRENA. (2022). *World energy transitions outlook 2022: 1.5c pathway.*

Jenkins, J. D., & Sepulveda, N. A. (2017). Enhanced decision support for a changing electricity landscape: the genx configurable electricity resource capacity expansion model.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260.

Kåberger, T. (2018). Progress of renewable electricity replacing fossil fuels. *Global Energy Interconnection*, *1*(1), 48–52.

Karschnia, B. (2022). *What's needed to close the skills gap in the power industry.* Retrieved from `https://www.forbes.com/sites/forbestechcouncil/2022/04/08/whats-needed-to-close-the-skills-gap-in-the-power-industry/?sh=3359353f138e`

Kelsall, C. C., Buznitsky, K., & Henry, A. (2021). Technoeconomic analysis of thermal energy grid storage using graphite and tin. *arXiv preprint arXiv:2106.07624*.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Konishi, S. (2014). *Introduction to multivariate analysis: Linear and nonlinear modeling*. CRC Press.

Kotsiantis, S. B., Zaharakis, I., Pintelas, P., et al. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, *160*(1), 3–24.

Kumar, K. R., & Kalavathi, M. S. (2018). Artificial intelligence based forecast models for predicting solar power generation. *Materials today: proceedings*, *5*(1), 796–802.

LaPotin, A., Schulte, K. L., Steiner, M. A., Buznitsky, K., Kelsall, C. C., Friedman, D. J., ... others (2022). Thermophotovoltaic efficiency of 40%. *Nature*, *604*(7905), 287–291.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436–444.

Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*.

Lopes, J. P., Hatziargyriou, N., Mutale, J., Djapic, P., & Jenkins, N. (2007). Integrating distributed generation into electric power systems: A review of drivers, challenges and opportunities. *Electric power systems research*, *77*(9), 1189–1203.

Mallapragada, D. S., Sepulveda, N. A., & Jenkins, J. D. (2020). Long-run system value of battery energy storage in future grids with increasing wind and solar generation. *Applied Energy*, *275*, 115390.

Manfren, M., Nastasi, B., Groppi, D., & Garcia, D. A. (2020). Open data and energy analytics-an analysis of essential information for energy system planning, design and operation. *Energy*, *213*, 118803.

March, J. G., & Olsen, J. P. (1976). Organizational choice under ambiguity. *Ambiguity and choice in organizations*, *2*, 10–23.

Mashlakov, A., Kuronen, T., Lensu, L., Kaarna, A., & Honkapuro, S. (2021). Assessing the performance of deep learning models for multivariate probabilistic energy forecasting. *Applied Energy*, *285*, 116405.

Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., ... others (2021). Climate change 2021: the physical science basis. *Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change*, *2*.

Mazzi, N., & Pinson, P. (2017). Wind power in electricity markets and the value of forecasting. In *Renewable energy forecasting* (pp. 259–278). Elsevier.

Medsker, L. R., & Jain, L. (2001). Recurrent neural networks. *Design and Applications*, *5*, 64–67.

Mehlum, E., Hischier, D., & Caine, M. (2021). *This is how ai will accelerate the energy transition.* Retrieved from `https://www.weforum.org/`

agenda/2021/09/this-is-how-ai-will-accelerate-the-energy-transition/

Merino, J., Gómez, I., Fraile-Ardanuy, J., Santos, M., Cortés, A., Jimeno, J., & Madina, C. (2021). Fostering der integration in the electricity markets. In *Distributed energy resources in local integrated energy systems* (pp. 175–205). Elsevier.

Mikalsen, K. Ø. (2019). Advancing unsupervised and weakly supervised learning with emphasis on data-driven healthcare.

Montavon, G., Lapuschkin, S., Binder, A., Samek, W., & Müller, K.-R. (2017). Explaining nonlinear classification decisions with deep taylor decomposition. *Pattern recognition*, *65*, 211–222.

Mosavi, A., Salimi, M., Faizollahzadeh Ardabili, S., Rabczuk, T., Shamshirband, S., & Varkonyi-Koczy, A. R. (2019). State of the art of machine learning models in energy systems, a systematic review. *Energies*, *12*(7), 1301.

Musavi, M. T., Ahmed, W., Chan, K. H., Faris, K. B., & Hummels, D. M. (1992). On the training of radial basis function classifiers. *Neural networks*, *5*(4), 595–603.

Orr, M. J., et al. (1996). *Introduction to radial basis function networks.* Technical Report, center for cognitive science, University of Edinburgh ....

Owerko, D., Gama, F., & Ribeiro, A. (2018). Predicting power outages using graph neural networks. In *2018 ieee global conference on signal and information processing (globalsip)* (pp. 743–747).

Palma, D., Varnajot, A., Dalen, K., Basaran, I. K., Brunette, C., Bystrowska, M., ... Ronge, T. A. (2019). Cruising the marginal ice zone: climate change and arctic tourism. *Polar Geography*, *42*(4), 215–235.

Panteli, M., & Mancarella, P. (2015). Influence of extreme weather and climate change on the resilience of power systems: Impacts and possible mitigation strategies. *Electric Power Systems Research*, *127*, 259–270.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... others (2011). Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, *12*, 2825–2830.

Perera, A., Nik, V. M., Chen, D., Scartezzini, J.-L., & Hong, T. (2020). Quantifying the impacts of climate change and extreme climate events on energy systems. *Nature Energy*, *5*(2), 150–159.

Qi, Z., Khorram, S., & Li, F. (2019). Visualizing deep networks by optimizing with integrated gradients. In *Cvpr workshops* (Vol. 2, pp. 1–4).

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1985). *Learning internal representations by error propagation* (Tech. Rep.). California Univ San Diego La Jolla Inst for Cognitive Science.

Salinas, D., Flunkert, V., Gasthaus, J., & Januschowski, T. (2020). Deepar: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, *36*(3), 1181–1191.

Sazli, M. H. (2006). A brief review of feed-forward neural networks. *Communications Faculty of Sciences University of Ankara Series A2-A3 Physical*

Sciences and Engineering, *50*(01).

Schittenkopf, C., Deco, G., & Brauer, W. (1997). Two strategies to avoid overfitting in feedforward networks. *Neural networks*, *10*(3), 505–516.

Selcuk, S. (2017). Predictive maintenance, its implementation and latest trends. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, *231*(9), 1670–1679.

Sepulveda, N. A., Jenkins, J. D., de Sisternes, F. J., & Lester, R. K. (2018). The role of firm low-carbon electricity resources in deep decarbonization of power generation. *Joule*, *2*(11), 2403–2420.

Sepulveda, N. A., Jenkins, J. D., Edington, A., Mallapragada, D. S., & Lester, R. K. (2021). The design space for long-duration energy storage in decarbonized power systems. *Nature Energy*, *6*(5), 506–516.

Shah, D., Patel, D., Adesara, J., Hingu, P., & Shah, M. (2021). Integrating machine learning and blockchain to develop a system to veto the forgeries and provide efficient results in education sector. *Visual Computing for Industry, Biomedicine, and Art*, *4*(1), 1–13.

Sharma, N., Sharma, P., Irwin, D., & Shenoy, P. (2011). Predicting solar generation from weather forecasts using machine learning. In *2011 ieee international conference on smart grid communications (smartgridcomm)* (pp. 528–533).

Sharma, S., Sharma, S., & Athaiya, A. (2017). Activation functions in neural networks. *Towards Data Sci*, *6*(12), 310–316.

Shrikumar, A., Greenside, P., & Kundaje, A. (2017). Learning important features through propagating activation differences. In *International conference on machine learning* (pp. 3145–3153).

Simonyan, K., Vedaldi, A., & Zisserman, A. (2013). Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034*.

Singh, H. V., Bocca, R., Gomez, P., Dahlke, S., & Bazilian, M. (2019). The energy transitions index: An analytic framework for understanding the evolving global energy system. *Energy Strategy Reviews*, *26*, 100382.

Skinner, T. (2018). *Is norway's tromso the new reykjavik?* Retrieved 2023-03-08, from `https://www.ft.com/content/a4fa6e78-ac3d-11e8-8253-48106866cd8a`

Springenberg, J. T., Dosovitskiy, A., Brox, T., & Riedmiller, M. (2014). Striving for simplicity: The all convolutional net. *arXiv preprint arXiv:1412.6806*.

Sundararajan, M., Taly, A., & Yan, Q. (2017). Axiomatic attribution for deep networks. In *International conference on machine learning* (pp. 3319–3328).

Suthaharan, S. (2016). Support vector machine. *Machine learning models and algorithms for big data classification: thinking with examples for effective learning*, 207–235.

Wuest, T., Weimer, D., Irgens, C., & Thoben, K.-D. (2016). Machine learning in manufacturing: advantages, challenges, and applications. *Production*

& Manufacturing Research, *4*(1), 23–45.

Yildiz, B., Bilbao, J. I., & Sproul, A. B. (2017). A review and analysis of regression and machine learning models on commercial building electricity load forecasting. *Renewable and Sustainable Energy Reviews*, *73*, 1104–1122.

Yin, J., Molini, A., & Porporato, A. (2020). Impacts of solar intermittency on future photovoltaic reliability. *Nature communications*, *11*(1), 4781.

Yin, W., Kann, K., Yu, M., & Schütze, H. (2017). Comparative study of cnn and rnn for natural language processing. *arXiv preprint arXiv:1702.01923*.

Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. In *Computer vision–eccv 2014: 13th european conference, zurich, switzerland, september 6-12, 2014, proceedings, part i 13* (pp. 818–833).

Zhang, Y., Wang, J., & Wang, X. (2014). Review on probabilistic forecasting of wind power generation. *Renewable and Sustainable Energy Reviews*, *32*, 255–270.

Zhou, S., Wang, Y., Zhou, Y., Clarke, L. E., & Edmonds, J. A. (2018). Roles of wind and solar energy in china's power sector: Implications of intermittency constraints. *Applied energy*, *213*, 22–30.

Ziegler, M. S., Mueller, J. M., Pereira, G. D., Song, J., Ferrara, M., Chiang, Y.-M., & Trancik, J. E. (2019). Storage requirements and costs of shaping renewable energy toward grid decarbonization. *Joule*, *3*(9), 2134–2153.

Åge Algerøy. (2023). *Kunstig intelligens kan hindra straumbrot*. Retrieved from `https://www.nrk.no/vestland/kunstig-intelligens-kan-hindra-straumbrot-1.16333843?fbclid=IwAR2G9aKbescSl7gkc10EksKqWY5DRP6RjQRWFCOyF8IcpUUfue-2YeBpoM8`