Doctoral theses at NTNU, 2021:385

Wei Li

Transparency and Explainability in Financial Data Science

Doctoral thesis

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of
Philosophiae Doctor
Faculty of Economics and Management
NTNU Business School

NTNU
Norwegian University of
Science and Technology

Wei Li

Transparency and Explainability in Financial Data Science

Thesis for the Degree of Philosophiae Doctor

Trondheim, December 2021

Norwegian University of Science and Technology
Faculty of Economics and Management
NTNU Business School

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Economics and Management
NTNU Business School

ISBN 978-82-326-6326-2 (printed ver.)
ISBN 978-82-326-5761-2 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

Doctoral theses at NTNU, 2021:385

Printed by NTNU Grafisk senter

Acknowledgments

This thesis is written for the degree of Philosophiae Doctor at the Norwegian University of Science and Technology (NTNU). The Ph.D. study acknowledges finance support by NTNU Research Grant: Financial challenges for the integration of short term electricity markets. The work acknowledges research support by COST Action "Fintech and Artificial Intelligence in Finance - Towards a transparent financial industry" (FinAI) CA19130. The research has been performed within the +CityxChange project under the Smart Cities and Communities topic that has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No. 824260.

I would like to express my sincere gratitude to Professor Florentina Paraschiv and Associate Professor Denis Mike Becker, my supervisors at NTNU Business School. Their guidance helped me in all research steps of this thesis. I am thankful for the trust I was given as well as for the understanding and unconditional support in difficult situations.

I gratefully acknowledge all the academic support obtained from NTNU Business School for my mobility at the University of St.Gallen, University of Cambridge, and the Humboldt University of Berlin. Special thanks to Arve Pettersen for his precious academic and moral support. I would also like to express my deepest gratitude to Professor Wolfgang Karl Härdle and the researchers in the International Research Training Group, who provided me valuable guidance throughout my academic exchange in Berlin.

Particular thanks go to my former supervisors, Georgios Sermpinis, Stein-Erik Fleten and Michael Phillips. Without their unconditional support and inspiring guidance, I would have never had the opportunity to pursue an academic career.

I am deeply indebted to my office mate: Ranik Raaen Wahlstrøm. Working with him is as much a honor as it is a pleasure. I would also like to thank all my former and current colleagues at NTNU Business School, making my Ph.D. study an enjoyment.

Many thanks to my family for all their love and encouragement, for raising me with a passion for science, and for supporting me in all my pursuits. I could not have completed this thesis without their support.

Summary

Financial data science has experienced rapid developments in recent years with the expansion of ever-growing data at an exponential rate. The proponents of data science argue that data science techniques will dominate and improve many domains of science in the next decades. However, several critics and concerns remain about its widespread adoption in the financial field due to the absence of transparency and explainability in the current generation of data science techniques. For instance, researchers' limited access to valuable data restrict the scientific developments and their benefits to both the academic community and industries. Additionally, while some state-of-the-art data science techniques, such as deep neural networks, have high prediction accuracy, they have been criticized for being black box methods that allow limited transparency into the decision process.

This thesis contributes to increasing transparency and explainability in financial data science by solving three types of research problems in energy and financial credit markets. First, my results shed more transparency on the intraday electricity trading by showing the impact of renewable energies on trader's strategies. In particular, I focus on the impact of wind and photovoltaic infeed on intraday electricity pricing. This study is particularly relevant to increasing transparency in intraday trading, since updates in weather forecasting errors are typically unavailable to researchers. Second, I employ state-of-the-art deep neural networks to price day-ahead electricity related to market coupling and use a post-hoc explainability technique to interpret prediction results. Third, I propose a data-driven explainable case-based reasoning method to predict financial credit risk, and show the relevance of its explainability in prediction results.

For intraday market traders, this thesis sheds light on how updated forecasts of renewable energies influence traders‘ behavior in the intraday trading. Moreover, it benefits intraday traders by proposing ways to model renewable energies forecasts that will further enhance existing econometric models for intraday electricity prices. Further, this thesis provides an efficient hybrid deep neural networks framework to predict day-ahead electricity prices under the consideration of market coupling for day-ahead electricity market participants. A post-hoc explainability technique is used to interpret the importance of the feature inputs, demand/supply variables, which offers more information and knowledge for cross-border market regulators and traders to move towards an integrated electricity market in Europe. Last, this thesis shows that financial institutions can benefit from the explainable case-based reasoning system to better serve their customers and reduce financial risk, in line with regulatory requirements. Compared with other machine learning methods, the proposed method provides superior prediction results of financial

risk and has a major relevance to the decision-making. This allows banks and other financial institutions to not only correctly map the probability of default for any borrower, but also to explain the underlying reason for default. In addition, results are highly relevant to borrowers, as it provides suggestions on how to improve their financial status to obtain new credit.

List of Articles

A) Modelling the Evolution of Wind and Solar Power Infeed Forecasts

B) Day-ahead Electricity Price Prediction Applying Hybrid Models of LSTM-based Deep Learning Methods and Feature Selection Algorithms under Consideration of Market Coupling

C) A Data-driven Evolutionary Case-based Reasoning Approach for Financial Risk Detection

Contents

1 Introduction **1**

2 Integration of Electricity Markets in Europe **5**

2.1 Day-ahead market 5

2.1.1 Day-ahead market price prediction in literature 6

2.1.2 Day-ahead market price prediction in the integrated market 9

2.2 Intraday market 10

2.2.1 Intraday market trading in the literature 11

2.2.2 Renewable energies and intraday markets 12

3 Explainable Artificial Intelligence **13**

3.1 Transparent machine learning models 14

3.1.1 Literature review 15

3.1.2 Case-based reasoning 16

3.2 Case-based reasoning for financial risk detection 18

3.2.1 Literature review 19

3.3 Post-hoc explainability techniques for machine learning models . 20

3.3.1 Literature review 20

3.3.2 Shapley Additive explanations 21

4 Research articles and contributions **23**

Article A: Modelling the evolution of wind and solar power infeed forecasts 23

Article B: Day-ahead electricity price prediction applying hybrid models of LSTM-based deep learning methods and feature selection algorithms under consideration of market coupling 24

Article C: A Data-driven Evolutionary Case-based Reasoning Approach for Financial Risk Detection 26

Bibliography **27**

Chapter 1

Introduction

In recent years, the quantity of financial data being generated has been expanding at an exponential rate. The ever-growing data boosts the digitization in the financial industry and the rapid development of financial data science, which has enabled methods and techniques such as advanced analytics, artificial intelligence, machine learning, and big data to penetrate and enhance the data analyses of financial institutions. Traditionally, financial analysis has extensively relied on the application of statistical inference and financial econometrics (Tsay 2005). However, certain drawbacks of financial econometrics have come under criticism. For instance, financial econometrics applies a limited number of models, such as regression models, which work in certain contexts but fail to deal with complex and high-dimensional data (Varian 2014, de Prado 2018). Moreover, financial econometrics is anchored toward the ideology of philosophical realism and statistical theory, rather than empirical forecasting, which has been detrimental to the ability of econometric models to produce reliable prediction results and constrained its practical applications (Summers 1991, Einav and Levin 2014, Varian 2014, Mullainathan and Spiess 2017).

With the advent of financial data science, researchers and practitioners combine statistics and computing in an effort to uncover patterns in datasets to expand the scope of financial econometrics and cope with its problems (Simonian and Fabozzi 2019). Practically, financial data science, considered as the expansion of the financial data ecosystem, represents an advancement over financial econometrics, and provides a wider range of innovations and robustnesses to practitioners and researchers when solving new and existing financial problems (Brooks et al. 2019, Khraisha 2020). As an application to credit risk management, financial data science allows a detailed analysis about debtors, which can help financial insti-

tutions to identify those in financial trouble and minimize their exposure to any potential default (Schmarzo 2013). For instance, Abakarim et al. (2018) apply a real-time deep neural networks approach to analyze loan applicant proposals and automatically make a decision for loan approval. Additionally, financial data science provides efficiency and cost-effectiveness for financial institutions to gain an understanding of how competitors perceive their products or identify customers' demand and design personalized products accordingly (Fang and Zhang 2016). In the study of Musto et al. (2015), a framework for a recommendation of asset allocation strategies is proposed, which combines case-based reasoning with a novel diversification strategy to assist financial advisers in the task of offering diverse and personalized investment portfolios. Moreover, the ever-increasing computational power allows researchers to experiment with an extremely large number of generated test subjects (Brooks et al. 2019). For example, Mclean and Pontiff (2016) and Jacobs and Müller (2020) study relevant explanatory variables of stock returns and explore the effect of predictability of these factors.

However, despite enormous optimism about the scope and variety of applications of financial data science, several critics and concerns remain about its widespread adoption. For instance, data has become a highly valuable commodity which in large part is owned and tightly controlled by a small number of technology companies and data brokers who have the advantage of formulating a strategy to harvest data from users and consumers. This can result in inequalities in access to data, which is problematic, since certain types of research could become restricted to a privileged few (Boyd and Crawford 2012). Inaccessible data can lead to limitations of other researchers conducting similar research projects and a reduction in innovation and transparency which would otherwise benefit both the research community and practitioners (Boyd and Crawford 2012). For example, the electricity intraday market lacks transparency as not all market participants and researchers have access to relevant variables such as weather updated forecasts (Kiesel and Paraschiv 2017). This is a drawback, because the inequality will lead to insufficient research on the intraday market mechanism.

From the industrial perspective, financial institutions always have an aversion towards advanced technology owing to its lack of transparency. For instance, machine learning methods have experienced rapid development in recent years due to their accurate prediction. However, some of them, such as neural networks, support vector machines, and tree ensembles, have been classified as black-box algorithms, since they are difficult to interpret and it is practically impossible to trace the logic involved in inference. Thus, these black-box methods cannot provide trust and confidence in their prediction due to the obfuscation for financial institutions, although they are more predictive than traditional models. Especially, under

the rule of the General Data Protection Regulation (GDPR) in Europe, decision-making based solely on automated processing is prohibited, while meaningful information about the logic involved should be carried on (The European Parliament and the Council of the European Union 2016). Additionally, the GDPR requires that any information or communication relating to the processing of personal data shall be easily accessible and easy to understand for the owner of these personal data. In such context, there is an imperative need for transparency and explainability in financial data science.

The purpose of this thesis is to contribute to the studies of transparency and explainability in financial data science by conducting research in three articles. In Article A, we investigate the transparency in electricity intraday market trading behaviors by exploring variables otherwise not directly accessible to traders and researchers. This sheds more transparency on the bidding strategies and shows the relevance of direct access to updated weather forecasts to electricity traders for correct adjustments of their positions. In Article B, we apply state-of-the-art long short-term memory (LSTM) based deep neural networks combined with feature selection methods to predict the Nordic system price, and used a theoretical game algorithm to detect the explainability of the proposed models in the prediction. In Article C, we propose a data-driven explainable case-based reasoning model to detect financial risk and show how to apply it to interpret prediction results.

The rest of this thesis is organized as follows. Chapter 2 discusses the theoretical background behind the first two research articles by giving an overview of the electricity market integration in Europe and motivates the need for transparency and explainability models required by the electricity markets. Chapter 3 describes the types of explainable artificial intelligence methods in the literature and their applications in this thesis. A summary of the three research articles included in this thesis, as well as their scientific contributions, are presented in Chapter 4.

Chapter 2

Integration of Electricity Markets in Europe

The current electricity wholesale market includes a series of sequential markets with different dynamics where bids are submitted and prices are determined: the day-ahead market, the continuous intraday market, and the balancing market. Day-ahead and intraday trading take place on market exchanges. In particular, market participants enter into bilateral contracts on purchases and sales of specific volumes of electricity at an agreed price and for delivery in an agreed period. The balancing energy market exists to enable grid operators to cost-effectively compensate for power and voltage fluctuations in the transmission grid.

Since the introduction of market coupling, the European Union is taking steps to improve the efficiency of the integration of the internal energy market and coupling of the European markets. Market Coupling aims to maximize the pan European social welfare by avoiding artificial splitting of the markets and contributing to the formation of the most relevant price signal for investment in cross-border transmission capacities. In particular, market coupling applies implicit auctions in which market participants do not individually receive allocations of cross-border capacity, but bid for the electricity on the cross-border exchange. The power exchanges take into account available cross-border capacity in the price calculation process, minimizing the price difference in different market areas.

2.1 Day-ahead market

The day-ahead market is the primary market for power trading where the largest volumes are traded. It is a market for contracts with the delivery of physical power

hour-by-hour the next day. For instance, participants make bids and offers to the Nord Pool trading system between 08:00 and 12:00 each day. Before 10:00 each day, the TSOs (Transmission System Operators) publish trading capacities for each bidding area. Prices for each hour of the following day are calculated on the basis of all the purchase and sell orders received and the transmission capacity available.

Under trade liberalization, the traditional vertically integrated power utilities are replaced with decentralized business entities whose targets are to maximize their profits. Consequently, a growing number of market participants are exposed to intense competition, and their need for suitable decision support models to increase margins and reduce risk has significantly increased (Bunn 2004). The availability of accurate day-ahead electricity price forecasts is vital for market participants to adjust production plans and perform effective bidding strategies to make an economic profit. Thus, accurate price prediction tools are essential for all electricity market participants for maximizing profits, mitigating risks, and stabilizing the grid under a liberalized and harmonized environment. Numerous research efforts have contributed to exploiting and developing advanced technologies for day-ahead energy price forecasting, aiming at highly accurate forecasting results (Weron 2014, Nowotarski and Weron 2018). In recent years, a considerable amount of literature has been devoted to electricity price forecasting models, which can be classified into five areas, each with its own strengths and weaknesses. These are multi-agent, fundamental, reduced-form, statistical, and computational intelligence (CI) models (Weron 2014). In general, the latter are state-of-the-art techniques. Compared with other traditional models, their performance superiority contributes to the prevalence of CI-based models in electricity price forecasting.

2.1.1 Day-ahead market price prediction in literature

In this section, I will shortly review some of the relevant literature that deals with electricity price forecasting. I will differentiate traditional methods on the one hand and the more recent CI approaches on the other.

Traditional models

One of the traditional methods for forecasting electricity prices is multi-agent models. They can be viewed as computerized systems that simulate the decisions and interactions of multiple, autonomous market participants. The agents are equipped with financial and other objectives; their behavior follows certain rules and mechanisms in order to achieve these objectives (Ventosa et al. 2005). Agent-based models are very flexible when it comes to the modeling of strategic decision-making and dynamics in the electricity market. However, the elements and relationships of such models must not only be theoretically founded a priori, but the

input parameters need to be calibrated in such a way that they correspond to the empirical reality (e.g., the number of market players, their trading strategies, and interaction mechanisms). Obviously, it is essential to collect the required information, and the unavailability of the majority of this information inevitably leads to potential modeling inaccuracy. To improve the performance of multi-agent models, researchers propose different hybrid approaches (Li et al. 2011, Zaman et al. 2017). In the study by Kiose and Voudouris (2015), an ACEWEM framework is proposed, which integrates the agent-based modeling paradigm with formal statistical methods, to simulate repeated power auctions.

The second category of traditional models comprise fundamental models which give insights into explicit formulations of the fundamental drivers of electricity prices, such as temperature, demand patterns, plant availability, and market mechanisms (Burger et al. 2007). The market price can be understood as an equilibrium price at the intersection of supply and demand. Despite numerous economic and physical factors incorporated in fundamental models (Gonzalez et al. 2012, Liebl 2013), unrealistic assumptions can result in the unreliability of the forecasting results. Furthermore, because of the limited availability of hourly data, concerning the fundamental drivers, fundamental models are not suitable for short-term price forecasting (Weron 2014).

The third category can be referred to as reduced-form models, which attempt to simulate the main dynamic characteristics of electricity prices. Typically, dynamic features include mean-reversion, jump-diffusion, and regime-switching (Islyaev and Date 2015). Such models provide a relatively simplified and tractable pattern of electricity prices, like spikes and volatility at a daily level. However, the literature report limited performance on day-ahead hourly prices when using reduced-form models (Bessec and Bouabdallah 2006, Weron and Misiorek 2008).

Finally, statistical models have been applied to the forecasting of electricity prices. Such models typically consist of an ARMA (autoregressive moving average), ARMAX (autoregressive moving average with exogenous variables), or GARCH (generalized autoregressive conditional heteroskedastic) component (Cuaresma et al. 2004, Conejo et al. 2005, Misiorek and Weron 2006, Koopman et al. 2007). These models incorporate lagged observations of electricity prices in addition to exogenous variables like consumption, production, and weather conditions. The prevalence of statistical models in scientific research can be attributed to their distinct interpretation of the results and their simple implementation. These models support short-term electricity price forecasting and the incorporation of related fundamental variables without requiring complex systems modeling. However, the efficiency and accuracy of such technical models are often criticized because their forecasting performance degrades when non-linearity or spikes are present, which

is typically the case in electricity price time series (Weron 2014). To deal with this drawback, nonlinear components have been exploited and integrated into statistical models (see Nogales et al. 2002, Jonsson et al. 2013, Gonzalez et al. 2018).

Computational intelligent models

Compared with traditional models, CI models can better handle complicated problems and compounded and dynamic systems. CI models comprise artificial neural networks, support vector machines, fuzzy systems, and evolutionary algorithms (Weron 2014). Recently, methods based on artificial neural networks (ANN) have received the most attention in the research. Their capability and flexibility to handle complex nonlinearities make them preferred candidates for predicting short-term electricity prices. For example, Catalao et al. (2007) propose a three-layered ANN that was trained by the Levenberg-Marquardt algorithm to predict short-term electricity prices in the electricity markets of mainland Spain and California. Keles et al. (2016) propose an ANN-based method for day-ahead electricity price forecasting and show that their approach has a better performance than the statistical benchmark models. In particular, to improve the performance of an ANN model, K-Nearest Neighbour (KNN) is used to select the most relevant input data, to reduce the computational efforts during training. One important finding from this research is that the electricity price was positively correlated with the price of the same hour one day before. In the study by Peter and Raglend (2017), an ANN model embedded with a particle swarm optimization (PSO) and Wavelet transformation (WT) approach is presented to predict the short-term market clearing price. PSO is used to optimize the weights of the ANN, while WT is utilized for decomposing electricity prices into a well-behaved series. The results from this research show the potential superiority of such hybrid models.

In recent years, deep neural networks (DNNs) have gradually entered scientific research related to electricity price forecasting. They are already regarded as the state-of-art approach in various other disciplines (Hinton et al. 2012, Bahdanau et al. 2014, Li et al. 2018). DNNs are the extension of the traditional neural networks and, due to their multiple layers, these networks can learn hierarchical and complex features of data much better than their shallow counterparts. In terms of their architecture, DNNs are often categorized into three main classes: Feed-forward Neural Networks (FNNs), Recurrent Neural Networks (RNNs), and Convolutional Neural Networks (CNNs). FNNs conduct the data flow unidirectionally from the input layer to the output layer and recognize the complex non-linear relationship between input and output. RNNs, also called feedback artificial neural networks, further allow information to move backwards. They are typically applied in sequence or time series data modeling by building extra mappings to hold relevant information from past inputs. Despite the superiority of RNNs in cap-

turing short-term dependencies, they are not capable of dealing with long-term dependencies due to the vanishing gradient problem (Bengio et al. 1994). The two most important variants of RNNs that overcome this problem are the long-short term memory (LSTM) and gated recurrent units (GRU) networks. CNNs have a different type of deep learning structure, including two special blocks, a convolution operation, and a pooling operation, commonly used to filter the features from data. Lago et al. (2018) broadly investigate the empirical performance of DNNs compared to traditional models for predicting day-ahead electricity prices. According to their results, DNNs outperform the statistical methods. Moreover, Chang et al. (2019) perform an empirical evaluation with the data in various day-ahead markets to demonstrate the state-of-art performance of LSTM models. Kuo and Huang (2018) detect the feasibility and practicality of electricity price forecasting by combining the CNN and LSTM models.

2.1.2 Day-ahead market price prediction in the integrated market

Over the last two decades, worldwide energy markets have experienced a transition towards deregulation and globalization (Weron 2006; 2014). This have resulted in more complex and integrated systems, making it harder to obtain accurate forecasts. A large number of explanatory variables from an ever-growing number of interconnected, neighboring power systems need to be considered when forecasting electricity prices. To the best of our knowledge, Article B of this thesis is the first study dedicated exclusively to exploring the influence of various features on electricity price forecasting by applying the state-of-the-art deep learning models under consideration of market coupling. In particular, we propose three hybrid architectures of LSTM-based DNN models: the two-step hybrid model, the auto-encoder hybrid model, and the two-stage hybrid model. Article B includes a case study which considers the system price forecasting of the Nord Pool day-ahead electricity market. We employ five feature selection algorithms for selecting feature variables derived from the markets listed on Nord Pool and their neighboring, interconnected markets. In Chapter 3, I will introduce the techniques to interpret the prediction results from black-box machine learning models.

The architecture of the LSTM network used in Article B is shown in Figure 2.1 and is defined by the following suit of equations (Graves 2013):

$$f_t = \sigma_g(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \tag{2.1}$$
$$i_t = \sigma_g(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \tag{2.2}$$
$$o_t = \sigma_g(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_{t-1} + b_o) \tag{2.3}$$
$$c_t = f_t \otimes c_{t-1} + i_t \otimes \sigma_h(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \tag{2.4}$$
$$h_t = o_t \otimes \sigma_h(c_t) \tag{2.5}$$

where f_t, i_t, o_t, c_t and h_t indicate the values of the forget gate state, input gate state, output gate state, memory cell and hidden state at time t in the sequence, respectively. σ_g and σ_h are sigmoid function and hyperbolic tangent function and $\otimes$ denotes the element-wise product. W and b are matrices/vectors with weights and biases which are coefficients to be estimated. Like all RNNs, the LSTM neural networks will process data sequentially. Hence, they take the form of a chain structure, as shown in Figure 2.2.

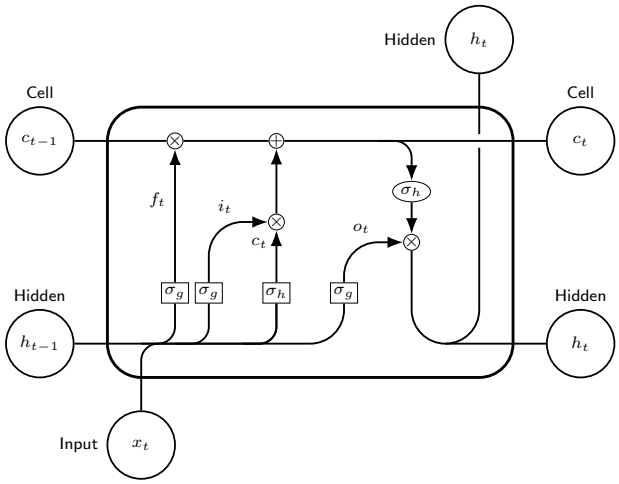


Figure 2.1: LSTM cell.

2.2 Intraday market

Intraday power trading refers to continuous buying and selling electricity at a power exchange. In the intraday market, contracts start to be continuously traded right after the closer of the day-ahead market and up to one hour before delivery. This allows a power plant operator who suddenly loses production in a single block to buy additional power from other participants on the market. In general, the electricity intraday market serves as a possibility to adjust the commitments from the

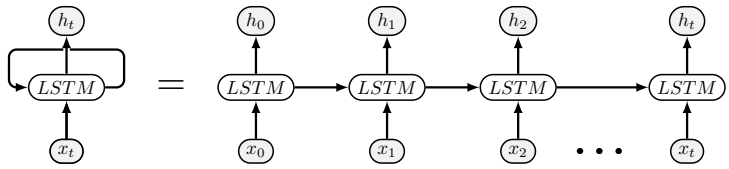


Figure 2.2: LSTM chain.

day-ahead market and reduce potential imbalance costs. Intraday markets start at different times in different countries. For example, in the German market, the intraday trading starts at 3 pm for hourly products and 4 pm for quarter-hourly, and ends 30 minutes before the delivery time (Kiesel and Paraschiv 2017). In the wake of the European Cross-Border Intraday (XBID) project, intraday trading becomes even more relevant. The aim of the XBID project is to create a single pan European cross zonal intraday market in Europe, which is essential for completing the European Internal Energy Market. With the rising share of intermittent renewable energies in the European generation, connecting intraday markets through cross-border trading is an increasingly important tool to provide more possibility for market parties to keep positions balanced.

2.2.1 Intraday market trading in the literature

While there is a rich number of price prediction models for the day-ahead market in the literature, similar research on intraday pricing lacks behind. Some recent research has paid attention to the intraday electricity price forecasting and trading (Monteiro et al. 2016, Ziel 2017, Kiesel and Paraschiv 2017, Kath and Ziel 2018, Uniejewski et al. 2019, Janke and Steinke 2019, Maciejowska et al. 2019, Narajewski and Ziel 2020a;b). Kiesel and Paraschiv (2017) provide an econometric model for 15-min intraday prices and show that these prices are asymmetrically influenced by intraday updated renewable forecast errors in a threshold regression analysis. Narajewski and Ziel (2020a) perform forecasting of intraday electricity prices in the German market and show that the intraday market is a weak-form efficient market. This outcome is consistent with Janke and Steinke (2019) who predict the quantiles of the German intraday price for the last three hours before delivery. Other studies of the German electricity market include Uniejewski et al. (2019), Narajewski and Ziel (2020b). Aïd et al. (2016) and Glas et al. (2020) develop optimal trading strategies for market participants who aim at marketing both renewable and conventional power on the intraday market. Gürtler and Paulsen (2018), Goodarzi et al. (2019) and Kulakov and Ziel (2021) detect the impact of renewable power generation and its forecast errors on intraday electricity prices.

The study of Kath and Ziel (2018) shows that the accurate prediction of 15-min intraday electricity price will generate economic benefits. Further they propose a buy-low-and-sell-high trading strategy based on the prediction to yield decent profits.

2.2.2 Renewable energies and intraday markets

In recent years, electricity generated by fossil fuels is replaced by the massive expansion of energy generated by renewable energy sources. This tendency has two direct influences on the electricity market. On the one hand, the increasing renewable power generation decreases the wholesale electricity price level because renewable power is increasingly cheaper than any new electricity capacity based on fossil fuels. On the other hand, it increases the volatility of electricity prices due to the fluctuating nature of the renewable energies generation profile, such as wind power and solar power. Thus, it becomes more challenging to obtain accurate electricity prices with the rapid development of intermittent renewable energies.

The German market is regarded as the world's first major renewable energy economy. Its renewable power generation increased from 38 TWh (7% of gross electricity production) in 2000 to 244 TWh (40% of gross electricity production) in 2019 (Federal Ministry for Economics Affairs and Energy 2020). In the light of the integration and interconnection of European electricity markets, the analysis of the renewables' influence on the German electricity market is essential for the local market but also for its integrated markets. Previous studies of Kiesel and Paraschiv (2017), Kremer et al. (2020a) and Kremer et al. (2020b) investigate the marginal effect of renewables forecasting errors on intraday electricity prices in the German market and show that negative (positive) forecast errors increase (decrease) the intraday price. However, the dataset of intraday evolution of solar and wind forecasting errors is not directly available to researchers ex-ante. The opacity of this vital information is an obstacle to understand trading behaviors in the electricity intraday market. For instance, Kremer et al. (2020b) and Narajewski and Ziel (2020b) observe that the closer to the delivery period, the greater the transaction frequency, especially in the last three hours in the German intraday market. However, the studies did not explain the reasons behind these observations. To the best of our knowledge, Article A of this thesis is the first to elaborate on uncovering hidden mechanisms in the bidding behavior by analyzing and modeling the wind and solar power infeed forecasts. The findings of Article A will benefit the intraday market participants as they can enhance existing econometric models for intraday electricity prices with the proposed accurate models for wind and solar power updated forecasting errors.

Chapter 3

Explainable Artificial Intelligence

In recent years, artificial intelligence (AI) has been developed rapidly and obtained notable achievements. AI relies on machine learning (ML) methods which have been advocated for their unprecedented levels of performance when learning to solve increasingly complicated problems and fulfill comprehensive tasks, which enable them to dominate the future development of the human society (West 2018). Further, the sophistication of AI methods enables them to be designed and deployed without human intervention, which explains why they have been criticized for being black box oracles that allow limited insight into decision factors (Lei et al. 2018). This includes deep neural networks (DNNs), which are complex neural network structures consisting of hundreds of layers and millions of parameters (Lei et al. 2018). The opaque decision process when using AI methods gives concern to their potential users which typically are reticent to adopt techniques that are not directly explainable, tractable and trustworthy (Zhu et al. 2018). Thus, there is an emerging need for understanding how such decisions are furnished by AI methods (Lipton 2016, Goodman and Flaxman 2017).

Particularly explainability of the working mechanism of a model is well appreciated by some decision-support systems where there is a preference to understand how the system produces a decision or recommendation, such as financial and medical systems (Moxey et al. 2010, Rai 2020). It is critical to develop and deploy trustworthy AI methods that meet financial and business objectives — from recommending products and content for customers, to personalizing user experience, to approving credit applications.

In addition, the current generation of black box AI algorithms are not suitable for use in regulated financial services (Bussmann et al. 2020). Explainability is one of

the main barriers AI algorithms face in regards to their widely practical implementation. To overcome this limitation, explainable AI (XAI), which provides reasons and transparency to make the functioning of AI clear and easy to understand, is in high demand (Bussmann et al. 2020). The targets of XAI are to provide an explainable and tractable reasoning process of decision-making and an understandable model mechanism. Typically, XAI pays more attention to the psychology of explanation and draws some insight from social science (Miller 2019). In contrast to black-box methods, which focus solely on performance and inevitably increase the opaqueness of the systems, XAI searches for a trade-off between the performance of a model and its transparency (Došilović et al. 2018). The improvement in the understanding of a system can lead to trust and confidence from human users.

XAI typically can be classified as transparent models and post-hoc explainability techniques. The duality derives from the two distinctive research directions (Guidotti et al. 2018). The former is related to the models which are interpretable by design. The latter relates to black box methods which can be explained by means of external XAI techniques.

3.1 Transparent machine learning models

Transparent ML models convey some degree of explainability by themselves. The levels of transparency in ML models can be evaluated based on three aspects: algorithmic transparency, decomposability and simulatability (Barredo Arrieta et al. 2020).

Algorithmic transparency is related to the ability of users to understand the process followed by the model to produce any given output from its input data (Barredo Arrieta et al. 2020). For instance, a linear model is regarded as transparent since it allows for relatively simple and interpretable inference, making it easy for the user to understand how the model will act in every situation it confronts (James et al. 2014). By contrast, the architectures of DNNs are considered to be opaque since the inference process cannot be fully observed and the obtained solution has to be approximated through heuristic optimization (Kawaguchi 2016, Datta et al. 2016), such as stochastic gradient descent. The explorable depth of models employing mathematical analysis and methods constrains their algorithmic transparency.

Decomposability is the ability to interpret individual parts of a model (Barredo Arrieta et al. 2020). For instance, additive models explicitly decompose a complex function into one-dimensional components, capturing non-linear relationships between individual features and the response, but retaining much of the intelligibility of linear models. By contrast, full complexity models, such as ensembles of trees, are more accurate on many datasets than additive models because they

model both nonlinearity and interaction, but too complex to provide any sufficient interpretation (Lou et al. 2012). In addition to the decomposability of the model itself, this characteristic requires every input feature to be readily interpretable. Typically, a transparent model is decomposable if every part of the model can be understandable by a human without the need for additional tools (Barredo Arrieta et al. 2020).

Simulatability indicates the ability of a ML model to be simulated or thought about strictly by a human (Barredo Arrieta et al. 2020). This means the complexity of models determines their interpretability in this aspect. In general, DNNs can provide accurate predictions, although the interpretation of such predictions obtained by DNNs is difficult (Yoshikawa and Iwata 2020). Meanwhile, linear models are more interpretable than complex non-linear models in spite of their predictive performance that would be inferior since real-world data is often intrinsically non-linear (Yoshikawa and Iwata 2020).

The explainability level of a transparent model can be evaluated as described above, namely algorithmic transparency, decomposability and simulatability.

3.1.1 Literature review

Typically, more complex models enjoy more flexibility than their simpler counterparts, allowing the complex models to have more accurate performance. In this context, it is unavoidable that the interpretability of the models decreases with an increased model complexity and performance. Figure 3.1 shows the relationship between the model accuracy and model transparency (Barredo Arrieta et al. 2020). From Figure 3.1, we can observe that DNNs are the most complex model with highest performance while rule-based learning and linear regression models are easy to the interpreter but with low prediction accuracy. The trade-off between model interpretability and performance is essential when developing a machine learning method (Barredo Arrieta et al. 2020). From this tentative representation, we can see that XAI has the potential and power to improve the common trade-off between model interpretability and performance.

Linear regression models, including their extensions like logistic regression models, clearly meet the characteristics of transparent models (algorithmic transparency, decomposability and simulatability) since they take the assumption of linear dependence between the explanatory variables and predictions. The usage of such models has been widely applied within financial fields for a long time and they are capable of explaining the results of the models to non-expert users (Ruppert 2004). In contrast, DNNs are welcomed by the academic community due to their huge ability to infer complex relations among variables although their complex

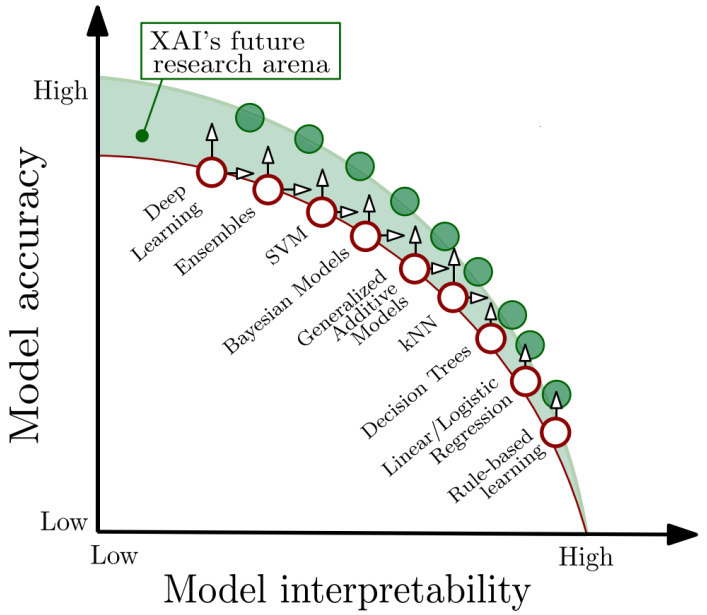


Figure 3.1: A representation of the area of improvement where the potential of XAI techniques and tools exists (Barredo Arrieta et al. 2020).

non-linear structure makes it hard to interpret their predictions (Pouyanfar et al. 2018). The fact that explainability of models is often a compulsory condition when used in practical forced the community to generate multiple explainability techniques (Pouyanfar et al. 2018), such as model simplification approaches, and feature relevance estimators (Thiagarajan et al. 2016, Che et al. 2017, Montavon et al. 2017).

3.1.2 Case-based reasoning

Case-based reasoning is a transparent machine learning method. It is the process of solving new problems based on the solutions of similar past cases (experiences). The principle is broadly based on how humans solve problems: solving new problems with past experiences in similar situations. The reasoning process typically consists of four steps (Aamodt and Plaza 1994) and the overview is given in Figure 3.2:

- Retrieve: Given a target problem, in the form of a case consisting of a problem and its corresponding attribute, the system searches among previously solved cases to find similar cases with solutions.

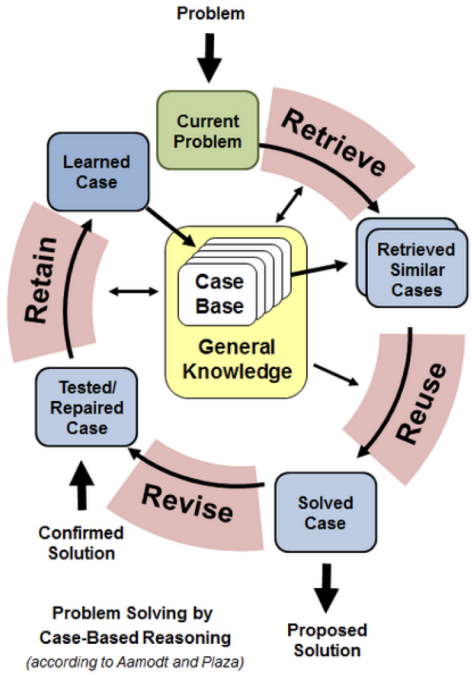


Figure 3.2: Overview of Case-based reasoning cycle (Aamodt and Plaza 1994). A case is generated from a problem (problem characterization) and the steps are completed in sequence to solve the problem.

- Reuse: Map the solutions from previous cases to the new problem. Adapting their solutions to generate a new solution to solve the new problem of the new case.
- Revise: Revise if the generated solution solved the target problem by applying the solution.
- Retain: If the new case has been solved based on the generated solution, the new case with the solution will be stored in the system.

Typically, the local-global principle is widely used in the attribute-based CBR system for case representation and similarity calculation (Richter and Weber 2013). Generally, the global similarity is typically measured by the square root of the weighted sum of all the local similarities. Given a query case Q and a case C from L-dimensional database (L features), a global similarity function $Sim(Q, C)$ to

calculate the similarity between Q and C can be described as follows:

$$Sim(Q,C) = \sqrt{\sum_{j=1}^{L} \boldsymbol{w_j} \times (sim_j(q_j, c_j))^2} \tag{3.1}$$

where, for the attribute j, sim_j is the local similarity function, q_j and c_j are attribute values from the case Q and C, respectively. $\boldsymbol{w_j}$ stands for the weight (global parameters) of the attribute j.

For the local (feature) similarity, asymmetrical polynomial functions are commonly used to measure the similarity of attribute-value (Bach and Althoff 2012). It can be represented as:

$$sim_j(q_j, c_j) = \begin{cases} \left(\frac{D_j - (c_j - q_j)}{D_j}\right)^{\boldsymbol{a_j}}, & \text{if } q_j \leq c_j \\ \left(\frac{D_j - (q_j - c_j)}{D_j}\right)^{\boldsymbol{b_j}}, & \text{if } q_j > c_j \end{cases} \tag{3.2}$$

where D_j stands for the difference between maximum and minimum value of attribute j in dataset. $\boldsymbol{a_j}$ and $\boldsymbol{b_j}$ are the degree (local parameters) of polynomial functions.

CBR level of explainability is analyzed as follows:

- Algorithmic transparency: the similarity measure cannot be fully observed and/or the number of variables is high. Thus, some mathematical and statistical algorithms would be used for the analysis of the model.
- Decomposability: the model comprises two equations: global similarity function and local similarity function. The similarity measure and the set of variables can be decomposed and analyzed separately. However, the amount of variables is too high and/or the similarity measure is too complex to be able to simulate the model completely.
- Simulatability: the complexity of the reasoning process matches human's native capabilities for simulation and solving new problems. The process is understandable by a human.

3.2 Case-based reasoning for financial risk detection

As financial markets grow increasingly complex, AI powered by DNNs can process large amounts of information and handle non-linear data, providing superior decision makings and yielding better investment returns compared to humans.

Nevertheless, important financial tasks as investment decision-making require explainability in the decision process. Similarly, it is also important to know why AI rejected a borrower's loan request and interpret why AI predicts one firm to be bankrupt and another not. However, few researchers have explored XAI in the financial field which is in high need.

Financial risks are potential losses associated with any form of financing, such as credit risk, operation risk, and business risk. Financial risk detection (FRD) is challenging as financial institutions not only require high accuracy but also an interpretable prediction process. The current generation of ML methods refers to the automated detection of meaningful patterns in data which have achieved considerably accurate results for FRD in recent years. However, it is very challenging to understand and explain the inner-workings of ML models, including how they accomplish their predictions and what variables are important (Honegger 2018). This leads to uncertainty as to whether black-box ML methods are suitable for solving problems in finance including detecting financial risk. The rule of GDPR implemented in 2018 further limited the application of black-box algorithms in the financial field (Voigt and Bussche 2017). Thus, the interpretability of ML methods is especially important for decision-makers who depend on analytics and data scientists for building sophisticated systems.

3.2.1 Literature review

FRD is typically a classification problem. In recent years, numerous ML methods have been developed and employed to improve the accuracy of FRD (Peng et al. 2011, Chen et al. 2011, Sermpinis et al. 2018, Hwang and Chu 2018, Lahmiri and Bekiros 2019). Chen et al. (2011) propose the Support Vector Machine (SVM) to predict the default risk of German firms and imply that the eight most important predictors related to bankruptcy for these German firms belong to the ratios of activity, profitability, liquidity, leverage, and the percentage of incremental inventories. Sermpinis et al. (2018) use the least absolute shrinkage and selection operator (LASSO) to predict market implied credit ratings and investigate the predictive determinants in relation to financial factors, market-driven indicators, and macroeconomic predictors. The results show LASSO models have superior predictive power and outperform the benchmark ordered probit models in all out-of-sample predictions. Lahmiri and Bekiros (2019) design an empirical study to assess the effectiveness of various ML methods for FRD and find a generalized regression neural topology outperforms multi-layer back-propagation networks, probabilistic neural networks, radial basis functions, and regression trees, as well as other advanced classifiers. They conclude that the utilization of advanced nonlinear classifiers based on big data methodologies can yield better bankruptcy forecasting accuracy compared to traditional ML methods. Overall, the majority of studies on

FRD achieve considerably accurate prediction results. However, no existing literature has discussed and explored the explainability of ML methods in financial decision makings.

In Article C of this thesis, a data-driven CBR model is employed for FRD. The aim of this article is to develop a way to automatically design the model to obtain accurate prediction results and explore its explainability in the process of solving the financial problem. The article detects five categories of financial risk for examining the predictability and explainability of the proposed case-based reasoning method, including credit card fraud, credit card default, credit default, bank churn, and financial distress.

3.3 Post-hoc explainability techniques for machine learning models

If ML models do not meet any of the criteria mentioned in section 3.1 to declare them transparent, an independent method is required to be designed and employed to explain their prediction process. The post-hoc explainability techniques are proposed and used for this purpose, which typically provides understandable information to analyze the model's predictions (Moradi and Samwald 2021).

In general, the post-hoc explainability techniques are categorized into two classes: those that are capable of being applied to any type of ML model and those that are devised for a specific ML model (Barredo Arrieta et al. 2020). The former is called model-agnostic techniques, such as Local Interpretable Model-Agnostic Explanations (LIME) (Ribeiro et al. 2016). The latter is referred to as model-specific techniques, such as DeepRED (rule extraction from deep neural networks) (Zilke et al. 2016).

3.3.1 Literature review

DNNs are considered to be the state-of-the-art ML methods in terms of their prediction accuracy. Numerous research has been conducted on the application of DNNs in the financial field (Cavalcante et al. 2016). Matsubara et al. (2018) propose a DNN generative model with information extracted from the news to predict stock prices. Their findings suggest that the proposed model performs better than SVM and multilayer perceptron (MLP) models. Futher, Fischer and Krauss (2018) use an LSTM-based DNN to predict stock prices and claim that LSTM can create an optimal trading system. Moreover, Wang et al. (2019) treat each kind of event, such as borrower's online click behavior, as a word, apply the Event2vec model to convert each kind of event into a vector, and use LSTM-based DNNs to predict the probability of default of borrowers. Chen et al. (2017) apply a DNN with

two layers for high-frequency forecasting stock prices and conclude that the proposed framework outperforms ARMA-GARCH and single-layer neural networks. Almahdi and Yang (2017) combine RNN and Reinforcement Learning to establish a portfolio of financial assets and find that the proposed management system responds to transaction cost effects efficiently and outperforms hedge fund benchmarks consistently.

However, DNNs lack explainability themselves, and thus, they need feature relevance techniques to interpret the results. Consequently, post-hoc local explanations and feature relevance techniques are increasingly the most adopted methods for explaining DNNs.

3.3.2 Shapley Additive explanations

SHAP (SHapley Additive exPlanations) are widely used to explain ML models, including DNN models (Lundberg and Lee 2017). This is a game theoretic approach to explain the output of ML models using Shapley values. Shapley values are a widely used in cooperative game theory, which distribute the total gains to the players, assuming that they all collaborate (Hart 2017). Let S be a coalition of players, then $v(S)$ describes the total expected sum of payoffs the members of S can obtain by cooperation. Given a coalitional game (v, N), the Shapley value of player i can be calculated as:

$$\varphi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!\,(n - |S| - 1)!}{n!} (v(S \cup \{i\}) - v(S)) \tag{3.3}$$

where N stands for the set of players. $S \subseteq N \setminus \{i\}$ means the sum extends over all subsets S of N not containing player i and n is the total number of players in each subset.

In Article B of this thesis, the Shapley value is used to assess the feature relevance relative to the output of ML models. In particular, we use the SHAP algorithm to interpret the impact of certain values of a given feature from the integrated electricity markets on the expected Nordic system price prediction.

Chapter 4

Research articles and contributions

This thesis consists of three articles. Below is a brief introduction as well as a description of the scientific contributions of each article.

Article A: Modelling the evolution of wind and solar power infeed forecasts

This article is co-authored with Prof. Dr. Florentina Paraschiv at NTNU. It is published in the *Journal of Commodity Market* (classified as ABS 3 and the impact factor is 2.721), and is available at `https://doi.org/10.1016/j.jcomm.2021.100189`.

In this article, we simulate and predict the evolution of wind and PV infeed forecasting errors over eight days preceding the start of a given quarter-hourly delivery period, updated in 15-min steps. In particular, we test comparatively the performance of three stochastic models, Ornstein-Uhlenbeck (OU) model, the Chan, Karolyi, Longstaff and Sanders (CKLS) model and Cox-Ingersoll-Ross (CIR) model, and a probabilistic model, Gaussian mixture model (GMM). In addition, we conduct robust statistical tests on the data set and show that the evolution of the weather infeed forecasts is a stationary process with a drift, but shows no volatility clustering. Our observations show the difference in updating the weather forecasts with respect to the time period left to the forecast (delivery) period. In particular, up to 8.5 hours (8 hours before the intraday market closes) in advance, adjustments are more frequently made, which increases the need for traders to adjust their positions in intraday trading. We delineate this time interval as the high-frequency-update period and the others as the low-frequency-update period. This finding is

consistent with the results in the study of Narajewski and Ziel (2020b), who show that the closer to the delivery period, the greater the transaction frequency. Furthermore, it explains that the liquidity of 15-min contracts rises significantly within the last trading hour prior to gate closure (Kremer et al. 2020b).

The scientific contribution of this article is threefold. First, we empirically analyze a novel and unique data set, and our proposed models break the ground for further applications to intraday pricing and optimization. This is important because intraday evolution of forecasting errors of 15-min updated solar and wind is essential to intraday market participants.

Second, we compare the in- and out-of-sample performance of four relevant methods. We find that the GMM yield a superior simulation performance versus the classical stochastic models when applied for simulating weather data. For the out-of-sample analysis, we find that the proposed models have different prediction performances, depending on whether the weather forecast updates follow a low- or a high-frequency pattern. The latter is recognizable the closer we come to the delivery period. In particular, the GMM performs better than the stochastic models during the low-frequency-update period. However, in the high-frequency-update period, when the time approaches the forecast period, stochastic models show superior performance. Thus, we recommend complimentary use of the proposed models, depending on the frequency in which forecast values are adjusted.

Third, the results of our simulation models for updated forecasting errors of solar and wind can be used as input to explore trading strategies and break the ground for enhanced pricing and optimization applications. In particular, accurate models for wind and PV updated forecasting errors can be used to enhance existing econometric models for intraday electricity prices. Simulations of wind and PV infeed forecasts are useful input to stochastic programming applications for optimal electricity production planning.

Article B: Day-ahead electricity price prediction applying hybrid models of LSTM-based deep learning methods and feature selection algorithms under consideration of market coupling

This article is co-authored with Dr. Denis Becker at NTNU. The article has been accepted by *Energy* (classified as level 2 in the Norwegian Scientific Index system and the impact factor is 7.147), and is available at `https://doi.org/10.1016/j.energy.2021.121543`.

In this article, we present the application of thirteen hybrid models of LSTM-based DNNs and feature selection algorithms for the day-ahead electricity price

prediction under consideration of market coupling. In particular, we introduce three hybrid LSTM-based architectures and investigate the performance of thirteen hybrid models. We find that the LSTM-based DNNs are overwhelmingly better than the benchmark statistical model (Nonlinear AutoRegressive Moving Average with eXogenous input (NARMAX) model). Also, we conclude that the different feature selection methods will lead to different feature selections. As input, diverse features will have a comparably significant impact on the performance of LSTM-based predictive models. Especially, the minimum redundancy maximum relevance algorithms, such as the RFE-SVR and Lasso regression feature selection methods, increase the performance of models (Radovic et al. 2017, Shao et al. 2017). In addition, we find the features production and consumption, as well as their prognosis in the Nordic and German markets, are prioritized by the feature selection method. Consequently, some implications can be provided for policymakers to improve cross-border trading in an integrated European power market. For instance, all trading capacity between the Nordic and German markets allocated to Nord Pool for implicit auction in the day-ahead price formation could result in a notable contribution to achieving better allocation of cross-border network capacity, such as the Nordic and Baltic bidding areas.

The scientific contribution of this article is threefold. First, three architectures of hybrid LSTM-based DNN models for electricity price forecasting are introduced. The obtained results in the empirical study show that the proposed models have considerably accurate prediction results for electricity prices. Further, we conclude that different feature selection algorithms yield divergent subsets of features, affecting the prediction accuracy of the proposed models. Our study is carried out using data from the Nord Pool market, but the generality of the proposed models ensures a possible application to other integrated markets, such as EPEX and OMIE.

Second, we provide an efficient way to utilize the ever-growing information from the electricity market integration for the Nordic EPF. This will benefit practitioners as these rely on accurate predictions (Zareipour et al. 2010, Kaminski 2013, Uniejewski et al. 2016).

Third, we employ a game theoretical SHAP approach to explore the relevance of various cross-border features in EPF. We illustrate the explainability of the SHAP method, and show the importance and impact of the different features from cross-border markets on EPF. The findings will benefit spot electricity traders and policymakers by better understanding the integrated market.

Article C: A Data-driven Evolutionary Case-based Reasoning Approach for Financial Risk Detection

This article is co-authored with Prof. Dr. Florentina Paraschiv at NTNU and Prof. Dr. Georgios Sermpinis at the University of Glasgow. The article has been submitted to *Quantitative Finance* (classified as ABS 3 and the impact factor is 2.222) at 07/07/2021.

In this article, we propose a data-driven evolutionary CBR system for FRD. The financial risk is typically associated with the potential loss in the financial field, such as credit risk, operation risk, and business risk. Compared to numerous black-box ML methods, which are shown to achieve considerably accurate results for FRD (Peng et al. 2011, Byanjankar et al. 2015, Sermpinis et al. 2018), the CBR system is an explainable AI model, being capable of explaining and justifying how it obtains its predictions. In addition, we find that the proposed CBR system has a good performance compared to other ML models.

The scientific contribution of this article is twofold. First, we propose a complete data-driven CBR system which is designed automatically without any required domain knowledge in the financial field. Our study solves the main drawback of a successfully developed CBR system which highly depends on prior experience and domain knowledge, which is challenging to acquire, even for experts. The generality of the proposed CBR system ensures a possible application to other decision-support systems where there is a preference to understand how the system makes the decision and produce the recommendation (Moxey et al. 2010).

Second, we introduce the four major goals of explanation in the CBR system and show the explainability of the CBR system in the empirical study. In particular, we show how to interpret the prediction results based on comparing similar cases and how to analyze the different importance of the features input. Moreover, we offer a way to calculate the posterior probability of the CBR system to justify prediction results. Furthermore, we show how to combine the data mining technique (clustering algorithm) and the CBR system to detect more information to enhance the financial system decision-making.

Bibliography

A. Aamodt and E. Plaza. Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Communications*, 7:39–59, 1994. doi: 10.3233/AIC-1994-7104. 1.

Y. Abakarim, M. Lahby, and A. Attioui. Towards an efficient real-time approach to loan credit approval using deep learning. In *2018 9th International Symposium on Signal, Image, Video and Communications (ISIVC)*, pages 306–313, 2018. doi: 10.1109/ISIVC.2018.8709173.

R. Aïd, P. Gruet, and H. Pham. An optimal trading problem in intraday electricity markets. *Mathematics and Financial Economics*, 10(1):49–85, Jan 2016. ISSN 1862-9660. doi: 10.1007/s11579-015-0150-8.

S. Almahdi and S. Y. Yang. An adaptive portfolio trading system: A risk-return portfolio optimization using recurrent reinforcement learning with expected maximum drawdown. *Expert Systems with Applications*, 87:267–279, 2017. ISSN 0957-4174. doi: https://doi.org/10.1016/j.eswa.2017.06.023.

K. Bach and K.-D. Althoff. Developing case-based reasoning applications using mycbr 3. In B. D. Agudo and I. Watson, editors, *Case-Based Reasoning Research and Development*, pages 17–31, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg. ISBN 978-3-642-32986-9.

D. Bahdanau, K. Cho, and Y. Bengio. Neural machine translation by jointly learning to align and translate. *arXiv:1409.0473 [cs.CL]*, 2014. Available from: `http://arxiv.org/abs/1409.0473`.

A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Her-

rera. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information Fusion*, 58:82–115, 2020. ISSN 1566-2535. doi: https://doi.org/10.1016/j.inffus.2019.12.012.

Y. Bengio, P. Simard, and P. Frasconi. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2):157–166, 1994. doi: https://doi.org/10.1109/72.279181.

M. Bessec and O. Bouabdallah. What causes the forecasting failure of Markov-Switching models? a Monte Carlo study. *Studies in Nonlinear Dynamics & Econometrics*, 9(2), 2006. doi: https://10.2202/1558-3708.1171.

D. Boyd and K. Crawford. Critical questions for big data. *Information, Communication & Society*, 15(5):662–679, 2012. doi: 10.1080/1369118X.2012.678878.

C. Brooks, A. G. F. Hoepner, D. McMillan, A. Vivian, and C. W. Simen. Financial data science: the birth of a new financial research paradigm complementing econometrics? *The European Journal of Finance*, 25(17):1627–1636, 2019. doi: 10.1080/1351847X.2019.1662822.

D. Bunn. *Modelling prices in competitive electricity markets.* Wiley, 2004.

M. Burger, G. Schindlmayr, and B. Graeber. *Managing Energy Risk: An Integrated View on Power and Other Energy Markets.* Wiley, 2007.

N. Bussmann, P. Giudici, D. Marinelli, and J. Papenbrock. Explainable ai in fintech risk management. *Frontiers in Artificial Intelligence*, 3:26, 2020. ISSN 2624-8212. doi: 10.3389/frai.2020.00026.

A. Byanjankar, M. Heikkilä, and J. Mezei. Predicting credit risk in peer-to-peer lending: A neural network approach. In *2015 IEEE Symposium Series on Computational Intelligence*, pages 719–725, 2015. doi: 10.1109/SSCI.2015.109.

J. Catalao, S. Mariano, V. Mendes, and L. Ferreira. Short-term electricity prices forecasting in a competitive market: A neural network approach. *Electric Power Systems Research*, 77(10):1297 – 1304, 2007. doi: https://doi.org/10.1016/j.epsr.2006.09.022.

R. C. Cavalcante, R. C. Brasileiro, V. L. Souza, J. P. Nobrega, and A. L. Oliveira. Computational intelligence and financial markets: A survey and future directions. *Expert Systems with Applications*, 55:194–211, 2016. ISSN 0957-4174. doi: https://doi.org/10.1016/j.eswa.2016.02.006.

Z. Chang, Y. Zhang, and W. Chen. Electricity price prediction based on hybrid model of adam optimized LSTM neural network and wavelet transform. *Energy*, 187:115804, 2019. ISSN 0360-5442. doi: https://doi.org/10.1016/j.energy.2019.07.134.

Z. Che, S. Purushotham, R. Khemani, and Y. Liu. Interpretable deep models for icu outcome prediction. *AMIA ... Annual Symposium proceedings. AMIA Symposium*, 2016:371–380, Feb 2017. ISSN 1942-597X. URL `https://pubmed.ncbi.nlm.nih.gov/28269832`.

H. Chen, K. Xiao, J. Sun, and S. Wu. A double-layer neural network framework for high-frequency forecasting. *ACM Trans. Manage. Inf. Syst.*, 7(4), Jan. 2017. ISSN 2158-656X. doi: 10.1145/3021380.

S. Chen, W. K. Härdle, and R. A. Moro. Modeling default risk with support vector machines. *Quantitative Finance*, 11(1):135–154, 2011. doi: 10.1080/14697680903410015.

A. J. Conejo, J. Contreras, R. Espínola, and M. A. Plazas. Forecasting electricity prices for a day-ahead pool-based electric energy market. *International Journal of Forecasting*, 21(3):435 – 462, 2005. doi: https://doi.org/10.1016/j.ijforecast.2004.12.005.

J. C. Cuaresma, J. Hlouskova, S. Kossmeier, and M. Obersteiner. Forecasting electricity spot-prices using linear univariate time-series models. *Applied Energy*, 77(1):87 – 106, 2004. doi: https://doi.org/10.1016/S0306-2619(03)00096-5.

A. Datta, S. Sen, and Y. Zick. Algorithmic transparency via quantitative input influence: Theory and experiments with learning systems. In *2016 IEEE Symposium on Security and Privacy (SP)*, pages 598–617, 2016. doi: 10.1109/SP.2016.42.

M. L. de Prado. *Advances in Financial Machine Learning*. Wiley Publishing, 1st edition, 2018. ISBN 1119482089.

F. K. Došilović, M. Brčić, and N. Hlupić. Explainable artificial intelligence: A survey. In *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pages 0210–0215, 2018. doi: 10.23919/MIPRO.2018.8400040.

L. Einav and J. Levin. Economics in the age of big data. *Science*, 346(6210), 2014. ISSN 0036-8075. doi: 10.1126/science.1243089.

B. Fang and P. Zhang. *Big Data in Finance*, pages 391–412. Springer International Publishing, Cham, 2016. ISBN 978-3-319-27763-9. doi: 10.1007/978-3-319-27763-9_11.

Federal Ministry for Economics Affairs and Energy. Energiedaten und -szenarien: Gesamtausgabe der energiedaten – datensammlung des bmwi, 2020. URL https://www.bmwi.de/Redaktion/EN/Artikel/Energy/energy-data.html.

T. Fischer and C. Krauss. Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2):654–669, 2018. ISSN 0377-2217. doi: https://doi.org/10.1016/j.ejor.2017.11.054.

S. Glas, R. Kiesel, S. Kolkmann, M. Kremer, N. Graf von Luckner, L. Ostmeier, K. Urban, and C. Weber. Intraday renewable electricity trading: advanced modeling and numerical optimal control. *Journal of Mathematics in Industry*, 10(1):3, Feb 2020. ISSN 2190-5983. doi: 10.1186/s13362-020-0071-x.

J. P. Gonzalez, A. M. S. M. S. Roque, and E. A. Pérez. Forecasting functional time series with a new Hilbertian ARMAX model: Application to electricity price forecasting. *IEEE Transactions on Power Systems*, 33(1):545–556, 2018. doi: https://doi.org/10.1109/TPWRS.2017.2700287.

V. Gonzalez, J. Contreras, and D. W. Bunn. Forecasting power prices using a hybrid fundamental-econometric model. *IEEE Transactions on Power Systems*, 27(1):363–372, Feb 2012. doi: https://doi.org/10.1109/TPWRS.2011.2167689.

S. Goodarzi, H. N. Perera, and D. Bunn. The impact of renewable energy forecast errors on imbalance volumes and electricity spot prices. *Energy Policy*, 134:110827, 2019. ISSN 0301-4215. doi: https://doi.org/10.1016/j.enpol.2019.06.035.

B. Goodman and S. Flaxman. European union regulations on algorithmic decision-making and a "right to explanation". *AI Magazine*, 38(3):50–57, Oct. 2017. doi: 10.1609/aimag.v38i3.2741.

A. Graves. Generating sequences with recurrent neural networks. *arXiv:1308.0850 [cs.NE]*, 2013. Available from: https://arxiv.org/abs/1308.0850.

R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi. A survey of methods for explaining black box models. *ACM Comput. Surv.*, 51 (5), Aug. 2018. ISSN 0360-0300. doi: 10.1145/3236009.

M. Gürtler and T. Paulsen. The effect of wind and solar power forecasts on day-ahead and intraday electricity prices in germany. *Energy Economics*, 75:150–162, 2018. ISSN 0140-9883. doi: https://doi.org/10.1016/j.eneco.2018.07.006.

S. Hart. *Shapley Value*, pages 1–5. Palgrave Macmillan UK, London, 2017. ISBN 978-1-349-95121-5. doi: 10.1057/978-1-349-95121-5_1369-2.

G. Hinton, L. Deng, D. Yu, G. E. Dahl, A. rahman Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, and B. Kingsbury. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 29:82 – 97, 2012. doi: https://doi.org/10.1109/MSP.2012.2205597.

M. Honegger. Shedding light on black box machine learning algorithms: Development of an axiomatic framework to assess the quality of methods that explain individual predictions, 2018.

R.-C. Hwang and C.-K. Chu. A logistic regression point of view toward loss given default distribution estimation. *Quantitative Finance*, 18(3):419–435, 2018. doi: 10.1080/14697688.2017.1310393.

S. Islyaev and P. Date. Electricity futures price models: Calibration and forecasting. *European Journal of Operational Research*, 247(1):144 – 154, 2015. doi: https://doi.org/10.1016/j.ejor.2015.05.063.

H. Jacobs and S. Müller. Anomalies across the globe: Once public, no longer existent? *Journal of Financial Economics*, 135(1):213–230, 2020. ISSN 0304-405X. doi: https://doi.org/10.1016/j.jfineco.2019.06.004.

G. James, D. Witten, T. Hastie, and R. Tibshirani. *An Introduction to Statistical Learning: With Applications in R*. Springer Publishing Company, Incorporated, 2014. ISBN 1461471370.

T. Janke and F. Steinke. Forecasting the price distribution of continuous intraday electricity trading. *Energies*, 12(22), 2019. doi: 10.3390/en12224262.

T. Jonsson, P. Pinson, H. A. Nielsen, H. Madsen, and T. S. Nielsen. Forecasting electricity spot prices accounting for wind power predictions. *IEEE Transactions on Sustainable Energy*, 4(1):210–218, 2013. doi: https://doi.org/10.1109/TSTE.2012.2212731.

V. Kaminski. *Energy markets*. Risk Book, 2013.

C. Kath and F. Ziel. The value of forecasts: Quantifying the economic gains of accurate quarter-hourly electricity price forecasts. *Energy Economics*, 76:411 – 423, 2018. ISSN 0140-9883. doi: https://doi.org/10.1016/j.eneco.2018.10.005.

K. Kawaguchi. Deep learning without poor local minima. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. URL https://proceedings.neurips.cc/paper/2016/file/f2fc990265c712c49d51a18a32b39f0c-Paper.pdf.

D. Keles, J. Scelle, F. Paraschiv, and W. Fichtner. Extended forecast methods for day-ahead electricity spot prices applying artificial neural networks. *Applied Energy*, 162:218 – 230, 2016. ISSN 0306-2619. doi: https://doi.org/10.1016/j.apenergy.2015.09.087.

T. Khraisha. A holistic approach to financial data science: Data, technology, and analytics. *The Journal of Financial Data Science*, 2020. ISSN 2405-9188. doi: 10.3905/jfds.2020.1.031.

R. Kiesel and F. Paraschiv. Econometric analysis of 15-minute intraday electricity prices. *Energy Economics*, 64:77–90, 2017. doi: https://doi.org/10.1016/j.eneco.2017.03.002.

D. Kiose and V. Voudouris. The acewem framework: An integrated agent-based and statistical modelling laboratory for repeated power auctions. *Expert Systems with Applications*, 42(5):2731 – 2748, 2015. doi: https://doi.org/10.1016/j.eswa.2014.11.024.

S. J. Koopman, M. Ooms, and M. A. Carnero. Periodic seasonal Reg-ARFIMA–GARCH models for daily electricity spot prices. *Journal of the American Statistical Association*, 102(477):16–27, 2007. doi: https://doi.org/10.1198/016214506000001022.

M. Kremer, R. Kiesel, and F. Paraschiv. Intraday electricity pricing of night contracts. *Energies*, 13(17), 2020a. doi: 10.3390/en13174501.

M. Kremer, R. Kiesel, and F. Paraschiv. A fundamental model for intraday electricity trading. *Philosophical Transactions of the Royal Society A, Forthcoming*, 2020b. doi: http://dx.doi.org/10.2139/ssrn.3489214.

S. Kulakov and F. Ziel. The impact of renewable energy forecasts on intraday electricity prices. *The Energy Journal*, Volume 10(1), Jan 2021. ISSN 1944-9089. doi: 10.5547/2160-5890.10.1.skul.

P.-H. Kuo and C.-J. Huang. An electricity price forecasting model by hybrid structured deep neural networks. *Sustainability*, 10(4), 2018. doi: https://doi.org/10.3390/su10041280.

J. Lago, F. D. Ridder, and B. D. Schutter. Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. *Applied Energy*, 221:386 – 405, 2018. doi: https://doi.org/10.1016/j.apenergy.2018.02.069.

S. Lahmiri and S. Bekiros. Can machine learning approaches predict corporate bankruptcy? evidence from a qualitative experimental design. *Quantitative Finance*, 19(9):1569–1577, 2019. doi: 10.1080/14697688.2019.1588468.

D. Lei, X. Chen, and J. Zhao. Opening the black box of deep learning. *CoRR*, abs/1805.08355, 2018. URL `http://arxiv.org/abs/1805.08355`.

G. Li, J. Shi, and X. Qu. Modeling methods for GenCo bidding strategy optimization in the liberalized electricity spot market–a state-of-the-art review. *Energy*, 36(8):4686 – 4700, 2011. ISSN 0360-5442. doi: https://doi.org/10.1016/j.energy.2011.06.015. PRES 2010.

L. Li, Z. Yuan, and Y. Gao. Maximization of energy absorption for a wave energy converter using the deep machine learning. *Energy*, 165:340 – 349, 2018. ISSN 0360-5442. doi: https://doi.org/10.1016/j.energy.2018.09.093.

D. Liebl. Modeling and forecasting electricity spot prices: A functional data perspective. *Ann. Appl. Stat.*, 7(3):1562–1592, 09 2013. doi: https://doi.org/10.1214/13-AOAS652.

Z. C. Lipton. The mythos of model interpretability. *CoRR*, abs/1606.03490, 2016. URL `http://arxiv.org/abs/1606.03490`.

Y. Lou, R. Caruana, and J. Gehrke. Intelligible models for classification and regression. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '12, page 150–158, New York, NY, USA, 2012. Association for Computing Machinery. ISBN 9781450314626. doi: 10.1145/2339530.2339556.

S. M. Lundberg and S.-I. Lee. A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 4768–4777, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.

K. Maciejowska, W. Nitka, and T. Weron. Day-ahead vs. intraday—forecasting the price spread to maximize economic benefits. *Energies*, 12(4), 2019. doi: 10.3390/en12040631.

T. Matsubara, R. AKITA, and K. UEHARA. Stock price prediction by deep neural generative model of news articles. *IEICE Transactions on Information and Systems*, E101.D(4):901–908, 2018. doi: 10.1587/transinf.2016IIP0016.

R. D. Mclean and J. Pontiff. Does academic research destroy stock return predictability? *The Journal of Finance*, 71(1):5–32, 2016. doi: https://doi.org/10.1111/jofi.12365.

T. Miller. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267:1–38, 2019. ISSN 0004-3702. doi: https://doi.org/10.1016/j.artint.2018.07.007.

S. Misiorek, A.and Trueck and R. Weron. Point and interval forecasting of spot electricity prices: Linear vs non-linear time series models. *Studies in Nonlinear Dynamics & Econometrics*, 10(3), 2006. doi: https://10.2202/1558-3708.1362.

G. Montavon, S. Lapuschkin, A. Binder, W. Samek, and K.-R. Müller. Explaining nonlinear classification decisions with deep taylor decomposition. *Pattern Recognition*, 65:211–222, 2017. ISSN 0031-3203. doi: https://doi.org/10.1016/j.patcog.2016.11.008.

C. Monteiro, I. J. Ramirez-Rosado, L. A. Fernandez-Jimenez, and P. Conde. Short-term price forecasting models based on artificial neural networks for intraday sessions in the iberian electricity market. *Energies*, 9(9), 2016. doi: 10.3390/en9090721.

M. Moradi and M. Samwald. Post-hoc explanation of black-box classifiers using confident itemsets. *Expert Systems with Applications*, 165:113941, 2021. ISSN 0957-4174. doi: https://doi.org/10.1016/j.eswa.2020.113941.

A. Moxey, J. Robertson, D. Newby, I. Hains, M. Williamson, and S.-A. Pearson. Computerized clinical decision support for prescribing: provision does not guarantee uptake. *Journal of the American Medical Informatics Association*, 17(1):25–33, 01 2010. ISSN 1067-5027. doi: 10.1197/jamia.M3170.

S. Mullainathan and J. Spiess. Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, 31(2):87–106, May 2017. doi: 10.1257/jep.31.2.87.

C. Musto, G. Semeraro, P. Lops, M. de Gemmis, and G. Lekkas. Personalized finance advisory through case-based recommender systems and diversification strategies. *Decision Support Systems*, 77:100–111, 2015. ISSN 0167-9236. doi: https://doi.org/10.1016/j.dss.2015.06.001.

M. Narajewski and F. Ziel. Econometric modelling and forecasting of intraday electricity prices. *Journal of Commodity Markets*, 19:100107, 2020a. ISSN 2405-8513. doi: https://doi.org/10.1016/j.jcomm.2019.100107.

M. Narajewski and F. Ziel. Ensemble forecasting for intraday electricity prices: Simulating trajectories. *Applied Energy*, 279:115801, 2020b. ISSN 0306-2619. doi: https://doi.org/10.1016/j.apenergy.2020.115801.

F. J. Nogales, J. Contreras, A. J. Conejo, and R. Espinola. Forecasting next-day electricity prices by time series models. *IEEE Transactions on Power Systems*, 17(2):342–348, 2002. doi: https://doi.org/10.1109/TPWRS.2002.1007902.

J. Nowotarski and R. Weron. Recent advances in electricity price forecasting: A review of probabilistic forecasting. *Renewable and Sustainable Energy Reviews*, 81:1548 – 1568, 2018. doi: https://doi.org/10.1016/j.rser.2017.05.234.

Y. Peng, G. Wang, G. Kou, and Y. Shi. An empirical study of classification algorithm evaluation for financial risk prediction. *Applied Soft Computing*, 11(2): 2906 – 2915, 2011. ISSN 1568-4946. doi: https://doi.org/10.1016/j.asoc.2010.11.028.

S. Peter and I. Raglend. Sequential wavelet-ANN with embedded ANN-PSO hybrid electricity price forecasting model for indian energy exchange. *Neural Comput & Applic*, 28:2277–2292, 2017. doi: https://doi.org/10.1007/s00521-015-2141-3.

S. Pouyanfar, S. Sadiq, Y. Yan, H. Tian, Y. Tao, M. P. Reyes, M.-L. Shyu, S.-C. Chen, and S. S. Iyengar. A survey on deep learning: Algorithms, techniques, and applications. *ACM Comput. Surv.*, 51(5), Sept. 2018. ISSN 0360-0300. doi: 10.1145/3234150.

M. Radovic, M. Ghalwash, N. Filipovic, and Z. Obradovic. Minimum redundancy maximum relevance feature selection approach for temporal gene expression data. *BMC Bioinformatics*, 18(1):9, Jan 2017. ISSN 1471-2105. doi: 10.1186/s12859-016-1423-9.

A. Rai. Explainable ai: from black box to glass box. *Journal of the Academy of Marketing Science*, 48(1):137–141, Jan 2020. ISSN 1552-7824. doi: 10.1007/s11747-019-00710-5.

M. T. Ribeiro, S. Singh, and C. Guestrin. "why should i trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 1135–1144, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450342322. doi: 10.1145/2939672.2939778.

M. M. Richter and R. O. Weber. *Case-Based Reasoning: A Textbook*. Springer Publishing Company, Incorporated, 2013. ISBN 364240166X.

D. Ruppert. *Statistics and Finance: An Introduction*. Springer Texts in Statistics. Springer, 2004. ISBN 9780387202709. URL https://books.google.no/books?id=DFJg_3PJ5ToC.

B. Schmarzo. *Big Data: Understanding How Data Powers Big Business*. Wiley, 2013. ISBN 9781118740002. URL https://books.google.no/books?id=WmukngEACAAJ.

G. Sermpinis, S. Tsoukas, and P. Zhang. Modelling market implied ratings using lasso variable selection techniques. *Journal of Empirical Finance*, 48:19–35, 2018. ISSN 0927-5398. doi: https://doi.org/10.1016/j.jempfin.2018.05.001.

Z. Shao, S. Yang, F. Gao, K. Zhou, and P. Lin. A new electricity price prediction strategy using mutual information-based svm-rfe classification. *Renewable and Sustainable Energy Reviews*, 70:330–341, 2017. ISSN 1364-0321. doi: https://doi.org/10.1016/j.rser.2016.11.155.

J. Simonian and F. J. Fabozzi. Triumph of the empiricists: The birth of financial data science. *The Journal of Financial Data Science*, 1(1):10–13, 2019. ISSN 2640-3943. doi: 10.3905/jfds.2019.1.010.

L. H. Summers. The scientific illusion in empirical macroeconomics. *The Scandinavian Journal of Economics*, 93(2):129–148, 1991. ISSN 03470520, 14679442. URL http://www.jstor.org/stable/3440321.

The European Parliament and the Council of the European Union. General Data Protection Regulation, 2016. URL http://data.europa.eu/eli/reg/2016/679/oj.

J. J. Thiagarajan, B. Kailkhura, P. Sattigeri, and K. N. Ramamurthy. Treeview: Peeking into deep neural networks via feature-space partitioning, 2016.

R. Tsay. *Analysis of financial time series*. Wiley series in probability and statistics. Wiley-Interscience, Hoboken, NJ, 2. ed. edition, 2005. ISBN 978-0-471-69074-0. URL http://gso.gbv.de/DB=2.1/CMD?ACT=SRCHA&SRT=YOP&IKT=1016&TRM=ppn+483463442&sourceid=fbw_bibsonomy.

B. Uniejewski, J. Nowotarski, and R. Weron. Automated variable selection and shrinkage for day-ahead electricity price forecasting. *Energies*, 9(8), 2016. ISSN 1996-1073. doi: https://doi.org/10.3390/en9080621.

B. Uniejewski, G. Marcjasz, and R. Weron. Understanding intraday electricity markets: Variable selection and very short-term price forecasting using lasso. *International Journal of Forecasting*, 35(4):1533 – 1547, 2019. ISSN 0169-2070. doi: https://doi.org/10.1016/j.ijforecast.2019.02.001.

H. R. Varian. Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2):3–28, May 2014. doi: 10.1257/jep.28.2.3.

M. Ventosa, A. Baillo, A. Ramos, and M. Rivie. Electricity market modeling trends. *Energy Policy*, 33(7):897 – 913, 2005. doi: https://doi.org/10.1016/j.enpol.2003.10.013.

P. Voigt and A. v. d. Bussche. *The EU General Data Protection Regulation (GDPR): A Practical Guide*. Springer Publishing Company, Incorporated, 1st edition, 2017. ISBN 3319579584.

C. Wang, D. Han, Q. Liu, and S. Luo. A deep learning approach for credit scoring of peer-to-peer lending using attention mechanism lstm. *IEEE Access*, 7:2161–2168, 2019. doi: 10.1109/ACCESS.2018.2887138.

R. Weron. *Modeling and forecasting electricity loads and prices: A statistical approach.* Wiley, 2006.

R. Weron. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, 30(4):1030 – 1081, 2014. ISSN 0169-2070. doi: https://doi.org/10.1016/j.ijforecast.2014.08.008.

R. Weron and A. Misiorek. Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models. *International Journal of Forecasting*, 24(4):744 – 763, 2008. doi: https://doi.org/10.1016/j.ijforecast.2008.08.004.

D. West. *The Future of Work: Robots, AI, and Automation*. Brookings Institution Press, 2018. ISBN 9780815732938. URL `https://books.google.no/books?id=W_zHtAEACAAJ`.

Y. Yoshikawa and T. Iwata. Neural generators of sparse local linear models for achieving both accuracy and interpretability. *CoRR*, abs/2003.06441, 2020. URL `https://arxiv.org/abs/2003.06441`.

F. Zaman, S. M. Elsayed, T. Ray, and R. A. Sarker. Co-evolutionary approach for strategic bidding in competitive electricity markets. *Applied Soft Computing*, 51: 1 – 22, 2017. ISSN 1568-4946. doi: https://doi.org/10.1016/j.asoc.2016.11.049.

H. Zareipour, C. A. Canizares, and K. Bhattacharya. Economic impact of electricity market price forecasting errors: A demand-side analysis. *IEEE Transactions on Power Systems*, 25(1):254–262, 2010. doi: https://doi.org/10.1109/TPWRS.2009.2030380.

J. Zhu, A. Liapis, S. Risi, R. Bidarra, and G. M. Youngblood. Explainable ai for designers: A human-centered perspective on mixed-initiative co-creation. In *2018 IEEE Conference on Computational Intelligence and Games (CIG)*, pages 1–8, 2018. doi: 10.1109/CIG.2018.8490433.

F. Ziel. Modeling the impact of wind and solar power forecasting errors on intraday electricity prices. In *2017 14th International Conference on the European Energy Market (EEM)*, pages 1–5, 2017.

J. R. Zilke, E. Loza Mencía, and F. Janssen. Deepred – rule extraction from deep neural networks. In T. Calders, M. Ceci, and D. Malerba, editors, *Discovery Science*, pages 457–473, Cham, 2016. Springer International Publishing. ISBN 978-3-319-46307-0.

Article A

Title:
Modelling the Evolution of Wind and Solar Power Infeed Forecasts

Authors:
Wei Li
Florentina Paraschiv

Reference:
Li W, Paraschiv F. Modelling the evolution of wind and solar power infeed forecasts. Journal of Commodity Markets: 100189, 2021. doi: `https://doi:org/10:1016/j:jcomm:2021:100189`.

Journal of Commodity Markets xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Commodity Markets

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

Modelling the evolution of wind and solar power infeed forecasts

Wei Li [a,*], Florentina Paraschiv [a,b]

[a] NTNU Business School, Norwegian University of Science and Technology, 7491, Trondheim, Norway
[b] University of St. Gallen, Institute for Operations Research and Computational Finance, Bodanstrasse 6, CH-9000, St. Gallen, Switzerland

ARTICLE INFO

Keywords:
Wind/photovoltaic forecasting errors
Intraday market
Gaussian mixture model (GMM)
Stochastic models

ABSTRACT

With the increasing integration of wind and photovoltaic power in the whole European power system, there is a longing for detecting how to trade energy in the ever-changing intraday market from electric power industries. Intraday trading becomes even more relevant in the wake of the European Cross-Border Intraday (XBID) project, which aims at integrating electricity trading across Europe. Therefore, optimal trading strategies to address forecast fluctuations in renewables output are growingly required to be designed. In this study, we model, simulate and predict the evolution of wind and PV infeed forecasting errors over eight days preceding the start of a given quarter-hourly delivery period and updated in 15-min steps. We test comparatively the performance of several stochastic and probabilistic models, and recommend their complementary use, depending on the frequency in which forecast values are adjusted. Since ex-ante updated forecasting errors of renewables infeed are usually not available to researchers, simulations based on our proposed models break the ground for further applications to intraday pricing and optimization.

1. Introduction

With a large amount of wind and photovoltaic power integrated into the power system in Germany in recent years, the intraday market traders recognize the importance of renewables forecasts. We analyze the eight days‘ evolution of 15-min updated forecasting errors of wind and photovoltaic (PV) for a specific quarter-hourly product traded in the intraday market, as given by weather data providers.[1] Typically, market participants in the intraday electricity trading aim at balancing out their positions after the closing of the day-ahead market. Given the volatile input from renewable energies, there is an increasing need for intraday trading. Since the updated weather (wind and solar) forecasted values can deviate significantly from the values published by the Transmission System Operators (TSOs) day-ahead, participants in the intraday market closely follow the evolution of updated forecasting errors. Updates are available every 15-min from the moment when the intraday market opens until shortly before the end. The intraday trading starts at 3 p.m. for hourly products and 4 p.m. for quarter-hourly, and ends 30 min before the delivery time in Germany (Kiesel and Paraschiv, 2017). Based on the updated information of renewable forecasts obtained from weather data providers on a 15-min basis, market participants adjust production schedules and price bids for specific quarter-hourly products (see the studies of Kiesel and Paraschiv (2017); Kremer et al. (2020a,b)).

The data set of intraday evolution of forecasting errors of solar and wind is not directly available to researchers ex-ante, but only ex-post, which is a drawback, given the market mechanism. Indeed, each trading day, market participants bid in the day-ahead and

* Corresponding author.
E-mail address: wei.n.li@ntnu.no (W. Li).

[1] source: EWE Trading GmbH https://www.ewe.com/en.

https://doi.org/10.1016/j.jcomm.2021.100189
Received 29 April 2020; Received in revised form 13 January 2021; Accepted 6 April 2021
Available online XXX

Please cite this article as: Li, W., Paraschiv, F., Modelling the evolution of wind and solar power infeed forecasts, Journal of Commodity Markets, https:/doi.org/10.1016/j.jcomm.2021.100189

intraday markets at EPEX. Price bids in both markets are based on expected values of supply/demand side explanatory variables, which are directly observable day-ahead from the TSOs. In the day-ahead market, electricity is traded for each hour of the next day, separately. However, due to the large forecasting errors in demand/supply variables and due to the increase of intermittent solar and wind infeed, there is a need to correct initial positions at a higher resolution. Thus, trading in the intraday market balances out excess demand/supply of electricity. An emerging challenge is how to secure adjustment capacity to respond to prediction errors and output fluctuations for renewable energy power generation.

Electricity is a special type of (non-storable) commodity, and electricity price forecasts are a fundamental input to energy companies' decision-making mechanisms, as discussed in the study of Weron (2006). An overview of forecasting modeling approaches for day-ahead prices is given in the study of Weron (2014). In recent years, the integration of smart grids and renewable energies has had the effect of increasing the uncertainty of future supply, demand, and prices (Nowotarski and Weron, 2018). This results in the growing interest of academics and practitioners to understand the probabilistic electricity price (and load) forecasting (Maciejowska et al., 2016; Liu et al., 2017). In the research article of Paraschiv et al. (2016b), the authors propose forecasting models for heavy-tailed electricity prices. The importance of renewable energies for accurate prediction of electricity price (extreme) quantiles is further explored in the studies of Hagfors et al. (2016a); Hagfors et al. (2016b); Frauendorfer et al. (2018) and Paraschiv et al. (2016a).

Recently, growing research focuses on the intraday electricity price forecasting (Monteiro et al., 2016; Ziel, 2017; Kath and Ziel, 2018; Uniejewski et al., 2019; Janke and Steinke, 2019; Maciejowska et al., 2019; Narajewski and Ziel, 2020a,b). Kath and Ziel (2018) employ an elastic net approach to forecast 15-min intraday prices and find that the intraday price in continuous trading is more challenging to predict than the auction price due to the flows of new information. Uniejewski et al. (2019) employ Lasso to address the problem of optimal variable choice for forecasting intraday prices in the German market and show that the most important explanatory variables are the most recently observed intraday and day-ahead prices. However, the study solely explores lagged price information, excluding demand/supply explanatory variables, such as renewable forecasts. Narajewski and Ziel (2020b) perform probabilistic forecasting of the hourly German intraday market price in the last 3 h of trading. Results indicate that the introduction of the XBID project reduces market volatility. Narajewski and Ziel (2020a) conduct an electricity price forecasting study in the German intraday market employing Lasso and elastic net techniques. The outcome shows that the intraday market is a weak-form efficient market. This is partially validated by Janke and Steinke (2019), who forecast the quantiles of the German intraday price for the last 3 h before delivery.

To the best of our knowledge, this is the first study that explores statistical properties of the evolution of renewables forecasts and tests suitable models for forecasting and simulation. Realistic simulation models are highly relevant to capture the typical volatility patterns of renewable forecast changes, which must be taken into account when building optimal intraday trading strategies. Given that ex-ante updated forecasting errors of wind and PV are generally not available to researchers, there is an enhanced need for realistic simulation models to explore trading strategies. Recent renewables-related econometric models are employed to simulate the energy output (Benth and Pircalabu, 2018; Benth and Ibrahim, 2017; Zhiwen et al., 2018) or to price electricity (Benth et al., 2007; Gibson and Schwartz, 1990). Research on intraday pricing incorporating ex-ante forecasts of wind and PV infeed is scarce. Previous studies (Kiesel and Paraschiv, 2017; Kremer et al., 2020a,b) investigate the marginal effect of renewables forecasting errors on intraday electricity prices and conclude that negative/positive forecast errors will increase/decrease the intraday price, respectively. Furthermore, Kremer et al. (2020a,b) show that 15-min intraday trading depends on the slope of the merit order curve. Thus, the authors show that market participants in the intraday market adjust their bids to the updated weather (wind/PV) forecasting errors asymmetrically. We found that the evolution of updated forecasts of wind and solar follows a mean reversion process, which is consistent with prior research. In a recent study, Kremer et al. (2020a) find clear evidence of mean reversion existing in the price formation mechanism of 15-min products. The study of Benth and Pircalabu (2018) shows a good fit of a mean reversion model to wind power production indexes, the model allowing for pricing wind power futures. The mean-reversion property is also explored in modeling photovoltaic power generation (Benth and Ibrahim, 2017; Boland, 2008). In practice, weather data providers make a rough approximation of the level of solar or wind infeed and forecasting errors are expected to oscillate around these expectations.

The contribution of this article is threefold. First, we empirically analyze a novel and unique data set of the evolution of 15-min updated wind and solar infeed forecasts, information essential to intraday market participants. Second, we test the in- and out-of-sample modeling performance of three relevant stochastic models (Ornstein Uhlenbeck (OU), Chan, Karolyi, Longstaff and Sanders (CKLS), and Cox-Ingersoll-Ross (CIR)) and one probabilistic model (the Gaussian Mixture Models, GMM). We conclude a superior in-sample simulation performance of GMM and recommend a complementary use of the tested models for out-of-sample forecasting. Third, the results of our simulation models for updated forecasting errors of solar and wind can be used as input to explore trading strategies and break the ground for enhanced pricing and optimization applications.

The rest of the paper is organized as follows: In Section 2, we describe the data set used by providing descriptive statistics. In Section 3, we introduce the technical specifications of the models used to describe the evolution of renewables infeed forecasts. In Section 4, we present estimation results. In Sections 5 and 6, we provide case studies to compare the performance of the proposed models in simulation and forecasting. Finally, Section 7 concludes.

2. Data and descriptive statistics

In this section, we provide a descriptive analysis of our data set of 15-min updated wind and solar infeed forecasts.

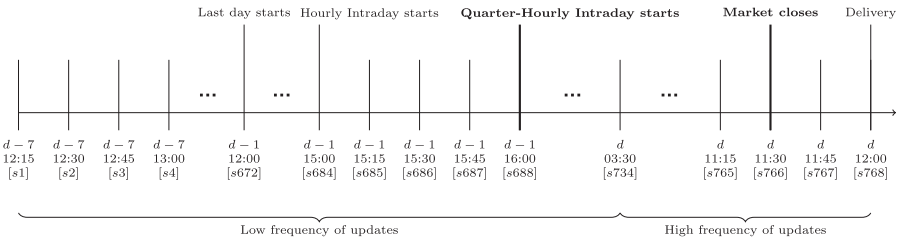


Fig. 1. The outline of timestamps for the wind/PV forecast updates for the H12Q1 product as forecast period. The evolution includes up to 8 days-ahead forecasts from day $d-7$ timestamp [$s1$] to the "delivery" day d timestamp [$s768$]. The low-frequency-update period is from step 1 [$s1$] to step 734 [$s734$], while the high-frequency-update period goes from step 735 [$s735$] to step 768 [$s768$].

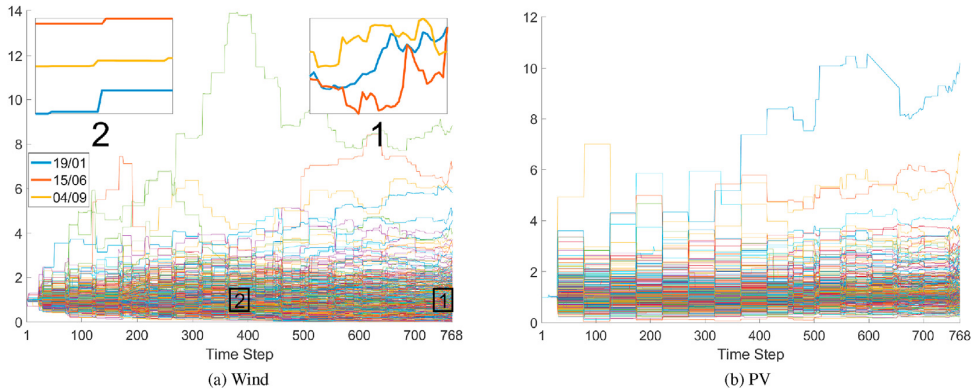


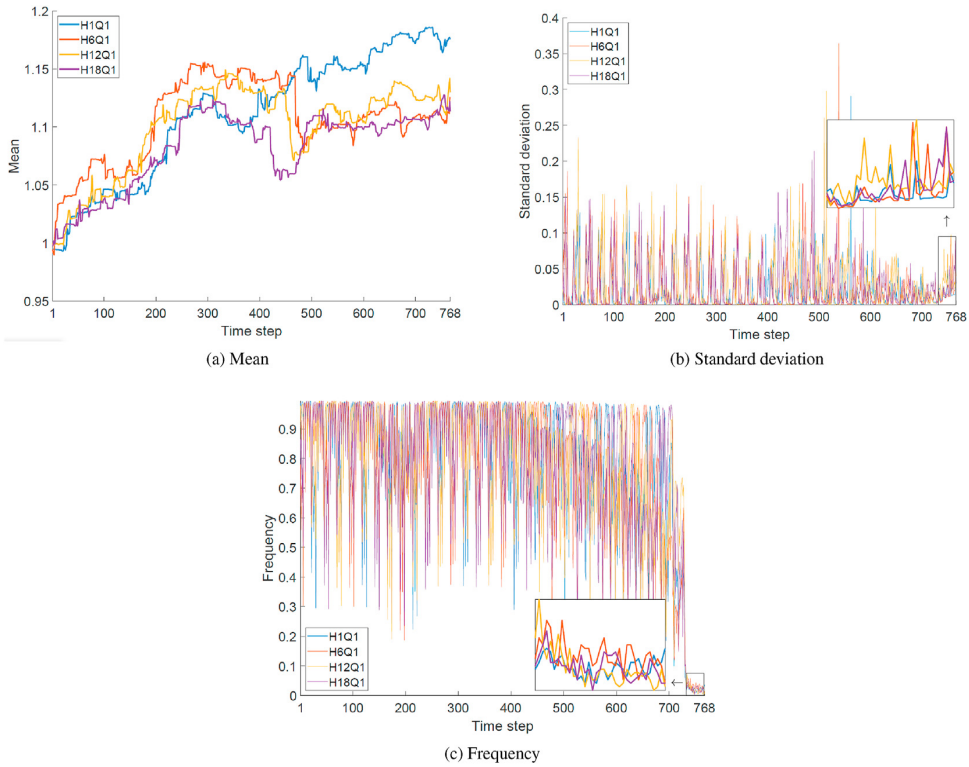
Fig. 2. The 8-day-ahead paths of updated forecasts for wind and PV for the fixed forecast (delivery) period represented by the quarter-hourly product H12Q1, observed each day in 2015. The zoom-in boxes 1, 2 show the difference in updating the weather forecasts with respect to the time period left to the forecast (delivery) period.

2.1. Data

We employ a unique data set of the evolution of updated forecasts for wind and PV infeed in 15-min frequency up to 8 days-ahead for a specific quarter-hourly product (forecast period) traded in the intraday electricity market each day between 01/01/2015 and 31/12/2015. Thus, for every quarter-hour forecast period, we observe 365 time series records of 768 weather forecasts each (up to 8 days × 24 hours × 4 quarter-hours-ahead). The data has been provided by the EWE Trading GmbH[2] and it relates to the intraday trading in the German electricity market at EPEX (see Epexspot (2019)). Information on wind/PV updated forecasts as ex-ante input to intraday trading is usually not available to researchers, but only ex-post. Our ex-ante/real time updates are therefore relevant to back-testing exercises/stress testing of trading strategies. As mentioned in the previous section, after the opening of the intraday market, traders receive updated forecasts of wind and PV every 15-min and adjust their trading strategies of quarter-hourly products accordingly. In Fig. 1, we outline the standard timestamps of updated wind/PV forecasts over 8 days prior to the selected forecast period: first quarter of hour 12 (H12Q1).

In 5% of the sample, our data set contains missing observations, which are handled by linear interpolation. To get a better comparative description of the evolution of forecasts across the 365 paths, we standardize the values to start at 1. In Fig. 2, we plot the paths of updated forecasts for wind and PV for the fixed forecast period represented by the quarter-hourly product H12Q1 (hour 12 quarter 1), observed each day in 2015. We observe that weather forecasts beyond 8.5-h-ahead are only adjusted step-wise, turning into continuous (frequent) adjustments closer to the start of the forecast period (see zoom-in sections of Fig. 2 for three selected paths, when the quarter-hourly forecast period is observed at 19/01, 15/06 and 04/09). This finding is in line with the common practice, where weather data providers run the forecasting models less frequently for updates of longer periods-ahead forecasts, but intensify their use closer to the forecast period. In Fig. 3(a), we show the mean of wind forecasts across the 365 paths for each time step. The drift in the evolution of wind forecasts is most pronounced for the forecast period H1Q1, which is not surprising, since wind is most intense during night hours. The frequency of no-updates in the wind forecasts evolution over time for several quarter-hourly forecast periods (observed at night, morning, noon and evening hours) is shown in Fig. 3(c). A frequency value of 1 indicates that for a specific quarter-hour forecast period across all 365 (days) paths at a given time step, no updates of wind forecasts are available. A

[2] https://www.ewe.com/en.



(a) Mean (b) Standard deviation

(c) Frequency

Fig. 3. (a) The mean of the wind infeed forecasts across all 365 paths at different time steps up to 8-day-ahead for the following forecast periods: H1Q1 (hour 1, quarter 1), H6Q1 (hour 6 quarter 1), H12Q1 (hour 12 quarter 1) and H18Q1 (hour 18 quarter 1). The zoom-in portions display the last 34 time steps (8.5 h before the delivery time). (b) The standard deviation of the wind infeed forecasts across all 365 paths at different time steps up to 8-day-ahead for the following forecast periods: H1Q1 (hour 1, quarter 1), H6Q1 (hour 6 quarter 1), H12Q1 (hour 12 quarter 1) and H18Q1 (hour 18 quarter 1). The zoom-in portions display the last 34 time steps (8.5 h before the delivery time). (c) The frequency (percentage) of no-update of the wind infeed forecasts across the 365 paths at specific 15-min interval time steps up to 8-day-ahead forecasts for the following forecast periods: H1Q1, H6Q1, H12Q1 and H18Q1 across the year 2015.

frequency value of 0.3 shows that in 30% of the days of one year, no updates of wind forecasts occurred at the given time step, for a specific quarter-hour forecast period. We observe that the frequency of updates increases closer to the forecast (delivery) period. As shown in Fig. 3(c), especially up to 8.5 h (8 h before intraday market closes) in advance, adjustments are more frequently made, which increases the need of traders to adjust their positions in the intraday trading. This is further shown in Fig. 3(b), where we observe an increase in the standard deviation of forecasts with 8 h before the market closure. This finding is in line with the results in the study of Narajewski and Ziel (2020b), who show that the closer to the delivery period (in our case corresponding to the "forecast period"), the greater the transaction frequency. Furthermore, Kremer et al. (2020a) show that the liquidity of 15-min contracts rises significantly within the last trading hour prior to gate closure: On average, roughly 68% of the total number of transactions are executed and about 74% of the total trading volume is transferred. As renewable power forecasts become more and more precise, traders in intraday market adjust their positions accordingly. The authors conclude that the 15-min periodicity in liquidity originates from newly arriving renewable forecast updates. Accurate forecast models for the latter are therefore essential for building realistic trading strategies.

2.2. Descriptive statistics

2.2.1. Stationarity and ARCH effects

Based on the visual inspection of Fig. 2, the evolution of updated forecasts of wind and PV show a drift over time. We apply the Augmented Dickey-Fuller test to detect whether the forecasts series have a unit root and thus follow a random walk. We test the null hypothesis of a random walk model (AR(1) coefficient = 1) with drift against the alternative model (AR(1) coefficient < 1) with drift. We replicated the unit root test to all forecast paths. The p-values and t-statistics are shown in Table 1, and the rejection rate of the null hypothesis at significance levels of 1%, 5%, 10% and 15% are shown in Table 2.

In Table 1, we observe that the mean of the p-values across the 365 paths is 0.0885, and their median is 0.0536. In Table 2, it is shown that the null hypothesis is rejected for up to 82.69% of the paths at 15% significance level. Thus, results show that the

Table 1
Results of the ADF test for unit root displayed for percentiles of the 365 paths of wind updated forecasts for the forecast period H12Q1 observed in 2015.

	mean	Std	min	25%	50%	75%	max
p-value	0.0885	0.1049	0.0001	0.0195	0.0536	0.1224	0.6758
t-statistic	−1.6369	0.6950	−3.6920	−2.0682	−1.6120	−1.1635	0.4560

The 25%, 50%, and 75% stand for the 25th, 50th, and 75th percentiles of the p-values or t-statistics across the 365 forecast paths, respectively.

Table 2
Rejection frequency of the ADF test for unit root for the 365 paths of wind updated forecasts for the forecast period H12Q1 observed each day in 2015, at different significance levels.

	1%	5%	10%	15%	Not rejected at 20%
Critical value	−2.331	−1.647	−1.283	−1.038	17.31%
Rejection rate	14.84%	47.53%	70.05%	82.69%	

Null hypothesis random walk model (AR(1) coefficient = 1) with drift against the alternative model (AR(1) coefficient < 1) with drift.

Table 3
Results of the Engle's ARCH test for all 365 wind forecast paths for the forecast period H12Q1 observed in 2015.

	mean	std	min	25%	50%	75%	max
p-value	0.6957	0.1973	0.0000	0.6184	0.7291	0.8327	0.9966
t-statistic	0.8565	6.0289	0.0000	0.0446	0.1200	0.2481	102.9716

The 25%, 50%, and 75% stand for the 25th, 50th, and 75th percentile, respectively, of p-values and t-tests across the 365 wind forecast paths having as forecast period H12Q1.

Table 4
Rejection frequency of the Engle's ARCH test for the 365 paths of wind updated forecasts, for forecast period H12Q1 observed in 2015, at different significance levels.

	1%	5%	10%	15%	Not rejected at 20%
Critical value	6.6349	3.8415	2.7055	2.0723	95.59%
Rejection rate	2.20%	3.03%	4.13%	4.41%	

Null hypothesis of no ARCH effects against the alternative that a series of residuals exhibits ARCH effects.

null hypothesis of unit root at common significance levels is rejected, indicating the presence of a drift in the stationary 15-min interval wind forecast updates series observed each day in 2015. Similar results have been obtained for the PV forecast paths and are available upon request.

We further employed the Engle's ARCH test to detect whether there are ARCH effects in the wind forecast time series. We conduct Engle's ARCH test (Engle, 1982) to test the null hypothesis that the series of wind forecast paths residuals exhibit no ARCH effects, against the alternative that autocorrelation exists in the squared residuals. Residual series are defined by $e = r$ - mean(r), where r is the return of time series (Engle, 1982). The test resulting p-values and t-statistics across all 365 wind updated forecast paths for the forecast period H12Q1 are shown in Table 3. The rejection rates of the null hypothesis at significance levels of 1%, 5%, 10% and 15% are shown in Table 4. Table 3 shows that the mean of the p-values across paths is 0.6957, and the median is 0.7291. Table 4 shows that in 95.59% of the cases the null hypothesis cannot be rejected at a significance level of 15%, indicating the absence of conditional heteroscedasticity in the variance process. In Fig. 4, we display the autocorrelation function and partial autocorrelation function for the squared residuals of the wind forecast paths for the forecast period H12Q1 observed at 08/03/2015. The sample ACF and PACF only have significant autocorrelation at lag 0, which implies that conditional heteroscedasticity does not prevail in the series.

Overall, we conclude that the evolution of the weather infeed forecasts is a stationary process with a drift, but shows no volatility clustering.

2.3. Distribution of forecasts across days of one year observed at each time step

For traders in the intraday market, it is important to get an overview of the distribution of forecasts of wind and PV at each time step to assess their volatility across the days of one year, which can further be used to adjust their bidding strategies. The closer we come to the delivery period, weather data providers let their forecasting models run more frequently and these updated forecasts incorporate a more realistic information which intraday traders typically follow closely. We therefore analyze the distribution of forecast returns across paths at each time step t along the 8 days of updated forecasts in 15-min frequency. For each step, we compute returns as shown in equation (1):

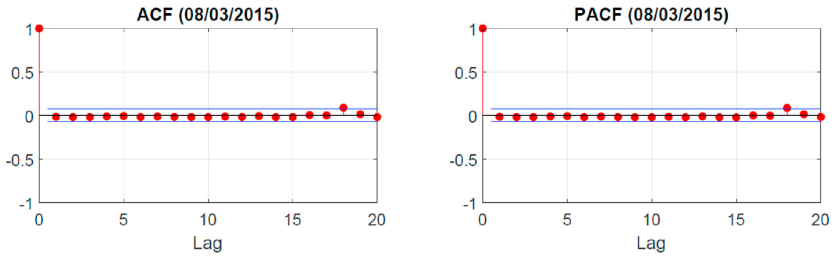


Fig. 4. The Autocorrelation Function and Partial Autocorrelation Function of the squared residual series of the wind forecast path observed at 08/03/2015 for the forecast period H12Q1.

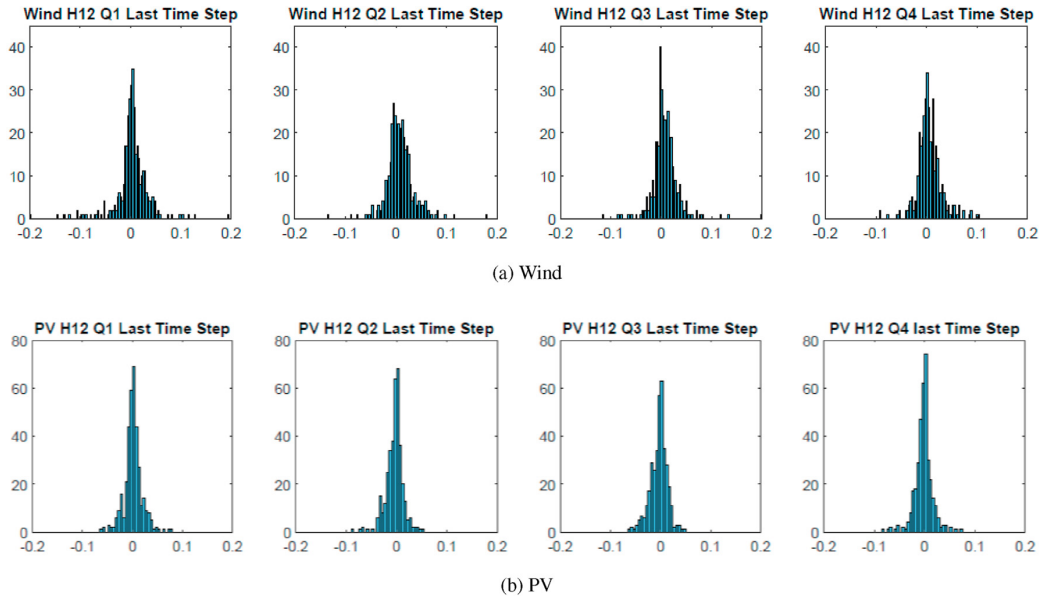


Fig. 5. The distribution of returns across all paths observed at the last time step for forecast period Hour 12, Quarters 1 to 4, for the wind and PV infeed forecasts, observed in 2015.

$$r_t^j = (X_t^j - X_{t-1}^j)/X_{t-1}^j, \quad for\ j = 1, 2, \ldots, 365 \tag{1}$$

where j represents the day when the (quarter-hourly) forecast period is observed. In Fig. 5, we show the histogram of the forecast returns across the 365 paths observed at the last time step before delivery time for each quarter-hourly product of hour 12 for wind and PV, respectively. We conclude that no quarter-hourly seasonality is observed. Data shows high kurtosis and skewed distribution. In addition, we show the histograms of wind and PV infeed forecasts in summer and winter (see Appendix: Figs. A.14 and A.15) and we find that there is no evidence supporting the existence of summer/winter seasonality in the renewables infeed forecasts. Thus, for the purpose of our analysis, we choose the peak quarter-hourly products (H12Q1), which show the highest trading volume in the intraday market.

We employed Jarque-Bera-Test to investigate whether the histograms of return forecasts observed across paths at a certain time step can be modeled by a normal distribution. We found that at a 5% significance level, the distribution is not normally distributed at all time steps. It should be noted that the p-values are less than 0.0001. This suggests that there is strong evidence that the return distributions are not normally distributed at a 5% significance level (results are available upon request).

We find that the weather forecasts series are stationary processes with drift, and show no volatility clustering along the path for a certain quarter-hourly product. Given the data properties, we tested three mean reversion models. In addition, to be able to simulate the distribution of forecasts across paths at given time steps, we test a probabilistic model where the entire data set is modeled as a set of simultaneous time series.

3. Methodology

In this paper, we will employ the Ornstein-Uhlenbeck (OU) model, the Chan, Karolyi, Longstaff and Sanders (CKLS) model, Cox-Ingersoll-Ross (CIR) model and the Gaussian mixture model, which are widely used to model energy time series.

3.1. Stochastic models

3.1.1. Ornstein-Uhlenbeck model

The mean reversion property often describes energy-related time series. The basic mean-reversion model is the Ornstein-Uhlenbeck (OU) process. The OU model and its modified versions were employed to model different phenomena in a broad range of fields such as economics, theoretical mathematics and physics (Gibson and Schwartz, 1990; Bibbona et al., 2008; Janczura et al., 2011). Specifically, the OU model is widely used in the field of energy, like commodity price and output simulation and prediction (Benth et al., 2007; Zárate-Miñano et al., 2013).

The Ornstein-Uhlenbeck process X_t is defined as a solution of the stochastic differential equation (SDE) of the following form (Glasserman, 2004):

$$dX_t = (\theta_1 + \theta_2 X_t)dt + \theta_3 dW_t \tag{2}$$

where $\theta_1 > 0$, $\theta_2 < 0$ and $\theta_3 > 0$ are parameters and W_t is the Wiener process. For the sake of convenience to interpret the solution of the SDE, its form can be transformed as follows:

$$dX(t) = \alpha(b - X(t))dt + \sigma dW(t) \tag{3}$$

where $\alpha = -\theta_2$, $b = \theta_1/\alpha$ and $\sigma = \theta_3$. α is the speed of reversion, b is the long-run mean and σ is the volatility. The general solution for the OU model is:

$$X(t) = e^{-\alpha t}X(0) + \alpha \int_0^t e^{-\alpha(t-s)}b(s)ds + \sigma \int_0^t e^{-\alpha(t-s)}dW(t) \tag{4}$$

To simulate X at time $0 = t_0 < t_1 < \ldots < t_n$, the process can be set as following equation in Euler scheme:

$$X(t_{i+1}) = X(t_i) + \alpha(b(t_i) - X(t_i))\Delta t + \sigma\sqrt{\Delta t}Z_{i+1} \tag{5}$$

where $\Delta t = t_{i+1} - t_i$ and Z_i with $i = 1, \ldots, n$ representing independent draws from $N(0, 1)$.

In the spacial case that $b(t) \equiv b$ and $\alpha > 0$, $X(t)$ converges in distribution to a normal distribution. The limiting distribution of $X(t)$ is a stationary distribution (Glasserman, 2004), with a limiting mean and variance given as:

$$\begin{aligned} E[X(t)] &= e^{-\alpha t}X(0) + (1 - e^{-\alpha t})b \to b, \quad as\ t \to \infty \\ Var[X(t)] &= \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha t}) \to \frac{\sigma^2}{2\alpha}, \quad as\ t \to \infty \end{aligned} \tag{6}$$

3.1.2. Cox-Ingersoll-Ross model and CKLS model

The volatility of the Ornstein-Uhlenbeck model is constant. In order to account for non-constant volatility, we apply the Cox-Ingersoll-Ross (CIR) model, which builds on the Ornstein-Uhlenbeck model. The volatility of generated forecasts from the CIR model is low for paths when forecasts approach zero but larger when the level of forecasts is high. Alain and Alexandre (2015) show that the CIR model has a good performance in forecasting wind speed on a short-term horizon. The CIR model is a special case of the CKLS model, which replaces the square root term with the general term, as shown in Equations (7) and (8). Thus, the CKLS model is our third tested stochastic model.

The Cox-Ingersoll-Ross model is defined as solutions of the stochastic differential equation of the following form:

$$dX_t = (\theta_1 + \theta_2 X_t)dt + \theta_3\sqrt{X_t}dW_t \tag{7}$$

where $\theta_1 > 0$, $\theta_2 < 0$ and $\theta_3 > 0$ are parameters and W_t is the Wiener process. The stochastic differential equation of CKLS Model is as follows:

$$dX_t = (\theta_1 + \theta_2 X_t)dt + \theta_3 X_t^{\theta_4}dW_t \tag{8}$$

where $\theta_1 > 0$, $\theta_2 < 0$, $\theta_3 > 0$ and $\theta_4 \neq 0$ are parameters and W_t is the Wiener process.

3.1.3. Parametric estimation of stochastic models

We fit the OU, CIR and CKLS stochastic models to the paths reflecting the evolution of updated forecasts of wind and PV and use the estimated model parameters to perform forecasting and simulation analyses. This task can be achieved by estimating parameters from discrete observations at the same time interval by maximum likelihood techniques. The well-known algorithms are the Euler method, the Ozaki method and the Shoji-Ozaki method. In this paper, we applied the Shoji-Ozaki method to estimate the parameters

of the aforementioned stochastic models. It has been shown that the performance of the Shoji-Ozaki method is much better than two other methods (Shoji and Ozaki, 1998). The stochastic differential equation can be considered as follows:

$$dX_t = f(t, X_t, \theta)dt + g(X_t, \theta)dW_t, \quad t \geq 0, X_0 = x_0, \tag{9}$$

as the transition density for the Shoji-Ozaki method is Gaussian, it can be written:

$$X_{t+\triangle t}|X_t = x \sim \mathcal{N}(A_{(t,x)}x, B^2_{(t,x)}), \tag{10}$$

$$A_{(t,x)} = 1 + \frac{f(t,x)}{xL_t}(e^{L_t \triangle t} - 1) + \frac{M_t}{xL_t^2}(e^{L_t \triangle t} - 1 - L_t \triangle t), \tag{11}$$

$$B_{(t,x)} = g(x)\sqrt{\frac{e^{2L_t \triangle t} - 1}{2L_t}}, \tag{12}$$

with:

$$L_t = \partial_x f(t,x) \quad and \quad M_t = \frac{g^2(x)}{2}\partial_{xx} f(t,x) + \partial_x f(t,x). \tag{13}$$

3.2. Gaussian mixture model

The probabilistic models build on probability distributions to simulate and predict the events. The basic idea is to build a probability space and choose random components that describe properties of the empirical distributions to simulate the stochastic process. The typical example is the Gaussian mixture model (GMM), which has been employed in studies that model wind forecast errors. In recent years, the GMM technique has been proven to be superior to classical time series models in fitting energy generation or loads of renewables (Singh et al., 2010; Valverde et al., 2012). Similar results are found in Zhiwen et al. (2018) who applied GMM and other distributions to generate scenarios for the wind power forecast errors showing superior explanatory power of the former.

A GMM is a probabilistic density model represented by a finite number of weighted Gaussian component densities (Reynolds, 2009). A GMM can be described as the following equation (Reynolds, 2009):

$$p(x|\lambda) = \sum_{i=1}^{M} w_i g\left(x|\mu_i, \sum_i\right) \tag{14}$$

where x is a continuous data vector, $w_i, i = 1, \ldots, M$, are the mixture weights, and $g(x|\mu_i, \sum_i)$, $i = 1, \ldots, M$, are the component Gaussian densities with mean vector μ_i and covariance matrix $\sum_i$. The mixture weights satisfy the constraint that $\sum_i^M w_i = 1$. λ is a notation collectively represented as:

$$\lambda = \{w_i, \mu_i, \sum_i\}, \ i = 1, \ldots, M. \tag{15}$$

The complete Gaussian mixture model is parameterized by mixture weights w_i from all component densities, the mean vectors μ_i and covariance matrices $\sum_i$. In this paper, GMM is adopted to represent the distribution of returns of the evolution of the wind and PV infeed updated forecasts observed across paths at each time step.

3.2.1. GMM parameter estimation

Obtaining the parameter set of the GMM is a typical parameter estimation problem. There are four steps to build a GMM. The procedure reads:

1. We need to decide about the number of components used for building GMM. In our case, we consider the number from 1 to 10.
2. Given a certain number of components, we apply the K-Means clustering method to find the initial center μ for each component. Mixture weights w and covariance matrices $\sum$ are initialized as well.
3. We iteratively adjust the initialized μ, w and $\sum$ to find the best fitted model.
4. After obtaining all the fitted models with different number of components, we select the best model from the collection via *AIC*.

4. Empirical results

4.1. Stochastic models

We implemented Shoji-Ozaki method to approximate the parameters of the stochastic models introduced in Section 3.[3] We estimate a SDE with Shoji-Ozaki method, using a discretization length of $\Delta t = \frac{1}{96}$ representing quarter-hourly data with the value

[3] The Sim.DiffProc is a R package which implements pseudo-maximum likelihood via the *fitsde*() function (Guidoum and Boukhetala, 2019).

Table 5
The estimated parameters for the path of wind quarter-hourly updated forecasts for the forecast period H12Q1 observed at 27/05/2015.

Model	θ 1	θ 2	θ 3	θ 4
OU	1.4144 (0.5792)	−1.6313 (0.6512)	0.1498 (0.0038)	
CKLS	0.82364 (0.4449)	−0.9533 (0.5267)	0.1935 (0.0078)	2.6610 (0.2515)
CIR	1.2498 (0.5462)	−1.4455 (0.6195)	0.1553 (0.0039)	

The value in (∗) is the standard error of parameter.

equaling 8 for frequency, representing the eight days of forecasts. The parameter estimates for the OU model, CIR model and CKLS model based on the path of wind quarter-hourly updated forecasts for the forecast period H12Q1 observed at 27/05/2015 are shown in Table 5.

We observe that the parameter signs of θ_1 for all calibrated stochastic models are positive, indicating a positive drift in the data and the θ_2 are negative, indicating the forecasts will revert to their long-run mean. This result confirms our previous descriptive statistics showing that the evolution of the forecasts is a mean-reverting process with a positive drift. The half-lives[4] of OU, CKLS and CIR models are 0.4249, 0.7271 and 0.4795, respectively. These numbers indicate that the evolution of the infeed forecasts need 10.19, 17.45 and 11.51 h, respectively, to reach the intermediate value between X(0) and the long run mean. The OU model requires the shortest time to converge. In addition, the scaled Wiener processes are the main difference among the three stochastic models. The volatility term θ_3 of the CIR model is multiplied by the term $\sqrt{X_t}$, and this eliminates the drawback of the OU model, a positive probability of receiving negative forecasts. When the forecasts approach zero, the volatility term $\theta_3\sqrt{X_t}$ approaches zero, canceling the randomness and implying that simulated forecasts will always stay positive. In contrast, high forecasts level correlates to high volatility, and this is a data property that can be observed in Fig. 2. As a generalization of the CIR model, CKLS targets to cover the observed forecasts more flexibly. In the CKLS model of this case study, the term $\theta_3\sqrt{X_t}$ is substituted by a more general term $\theta_3 {X_t}^{\theta_4}$ $\theta_4 = 2.6610$ indicates that the level of forecasts of CKLS model has more power of influence on the volatility term than CIR. Meanwhile, the flexibility of the volatility term of CKLS model leads to the possibility of creating negative values when the generated forecasts are near zero. In such a case, we will continue to generate new scenarios to substitute the ones with negative values. In the end, we ensure that all values of the generated scenarios of infeed forecasts are positive. In addition, the Ornstein–Uhlenbeck process can also be considered as the continuous-time analogue of the discrete-time AR(1) process ($x_t = c + \varphi x_{t-1} + w_t$). It is worthy to notice that $\varphi = \exp(\theta_2 \Delta t)$ based on Equation (4). In the current case, $\varphi = 0.9831$, which indicates a very strong autocorrelation in the series of forecast changes. The distribution of the φ parameters over all paths of the dataset in 2015 are shown in Appendix Fig. A.16. The high values of φ across paths confirm a strong autocorrelation in renewables infeed forecast changes.

Fig. 6 shows the parameter distributions for the OU, CKLS, and CIR models for Wind H12Q1, respectively. We can observe that the parameter distributions θ_1, θ_2, and θ_3 are either right or left skewed. The θ_4 distribution of CKLS model is symmetric but not normally distributed. Besides, we can also observe that several outliers fall far outside of the mass of parameter distributions. They are used to simulate abnormal evolution of renewable infeed forecasts. This is of high relevance to traders who need to balance out extreme/unexpected renewable output in the intrady market. The statistics of the described parameters above are shown in Table 6. We can find that, except the standard deviation of θ_2^{CIR} is slightly larger than the others, there is no obvious difference of θ_1 and θ_2 among the three models. However, the standard deviation of θ_3^{CIR} is noticeably lower than the other two, indicating that the CIR-based simulated values will have relatively stable volatility.

4.2. Gaussian mixture models

Many commercial software tools, like MATLAB, provide reliable solvers for K-Means clustering and GMM parameter estimation. If the model number of GMM is set to 1, the GMM will degenerate into being a Gaussian distribution. The Akaike information criterion (AIC) is an estimator of the relative quality of statistical models for a given set of data. AIC provides a means for model selection. $AIC = 2k - 2ln(L)$, k is the number of estimated parameters in the model, L is the maximum value of the likelihood function.

In this paper, we employ GMM to model the non-Gaussian distribution of renewable infeed forecast values across the 365 days at each forecasting step along the 8 days. The joint distribution of the forecasts return at each time step is modeled by the GMM with multivariate components. Thereafter, 767 joint distributions of the forecast returns for wind taking H12Q1 as forecast period are computed. The constructed GMM distributions of the last time step with the histogram of forecast returns are shown in Fig. 7. It can be seen that the joint distribution constructed by the GMM is consistent with the histograms through different numbers of components. Specifically, a careful look at the density histogram of the last time step suggests that the return values of the forecasts can be captured by the mixture components 1, 2 and 3 together. The peak of return values can be represented by component 1. The middle part of the distribution can be represented by the weighted combination of components 1 and 2. Since outliers exist in the return values, it is necessary to apply component 3 to capture the heavy tails.

Then, we assess the comparative performance of the GMM to the following distributions: normal-, stable-, beta-, lognormal-, gamma- and t-Location-Scale (generalized Student's t-distribution) distributions. We shift the data to a positive value before we

[4] The half-life (H) of a variable X is defined as the time for the expected value of X(t) to reach the middle value between X(0) and the long run mean. It is used to measure the strength of a mean-reversion process. The formula of the half-life is $H = \frac{\ln(2)}{|\theta_2|}$.

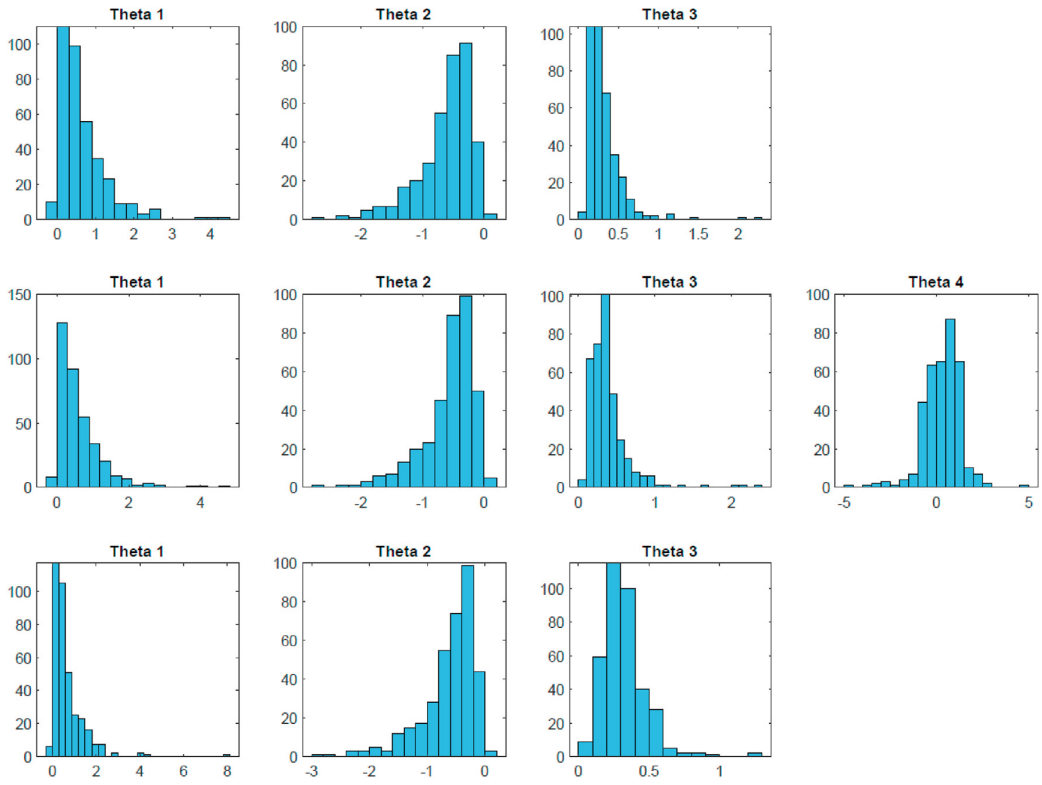


Fig. 6. Parameter distributions of Wind H12Q1 for OU model, CKLS model and CIR model across all 365 forecast paths.

Table 6
Statistics over the parameters of stochastic models. The 25%, 50%, and 75% stand for 25th, 50th, and 75th percentile, respectively.

	mean	std	min	25%	50%	75%	max
θ_1^{OU}	0.6489	0.6332	−0.1117	0.2113	0.4966	0.8736	4.4662
θ_2^{OU}	−0.6108	0.4448	−2.6759	−0.7898	−0.4944	−0.3140	0.0888
θ_3^{OU}	0.3221	0.2249	0.0720	0.1889	0.2714	0.3795	2.2061
θ_1^{CKLS}	0.6006	0.6237	−0.0859	0.1790	0.4254	0.8388	4.5200
θ_2^{CKLS}	−0.5623	0.4219	−2.6285	−0.6938	−0.4503	−0.2928	0.0855
θ_3^{CKLS}	0.3796	0.2605	0.0751	0.2343	0.3290	0.4452	2.3214
θ_4^{CKLS}	0.3290	0.9734	−4.5555	−0.2818	0.4313	0.9753	4.8111
θ_1^{CIR}	0.6591	0.7410	−0.0588	0.1964	0.4580	0.8651	7.8687
θ_2^{CIR}	−0.6139	0.4753	−2.8751	−0.7879	−0.4813	−0.3074	0.0416
θ_3^{CIR}	0.3165	0.1531	0.0717	0.2184	0.2992	0.3840	1.2617

apply beta-, lognormal- and gamma distributions. The fitted distributions along with GMM at the last time step are shown in Fig. 8. The Chi-square statistics are applied for a direct comparison of the goodness of fit. χ^2-statistics are measures to obtain the goodness of fit. Table 7 displays the χ^2 for various distributions. A better fit performance has a lower χ^2 value. The smallest value of χ^2 corresponds to the GMM method which confirms its best-fit performance among all considered distributions.

4.2.1. Improvement of performance

At some time steps, the goodness of fit performance of GMM is not good enough. One main reason is that the return data consist of many zero values, implying that the adjustment of the forecast does not occur at the current time step. To solve this issue, we employ a Bernoulli distribution to imitate a determination of "change", like setting a threshold. Bernoulli distribution is the discrete probability distribution of a random variable which takes the value 1 with probability p and the value 0 with probability $q = 1 - p$. It is also called 0–1 distribution. Following the application of Bernoulli distribution, GMM reflects the adjusted distribution of

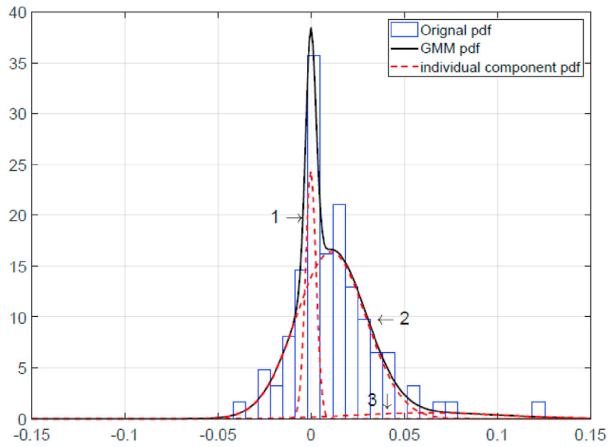


Fig. 7. The GMM applied to the wind updated forecasts for forecast period H12Q1 across all 365 paths observed at the last time step.

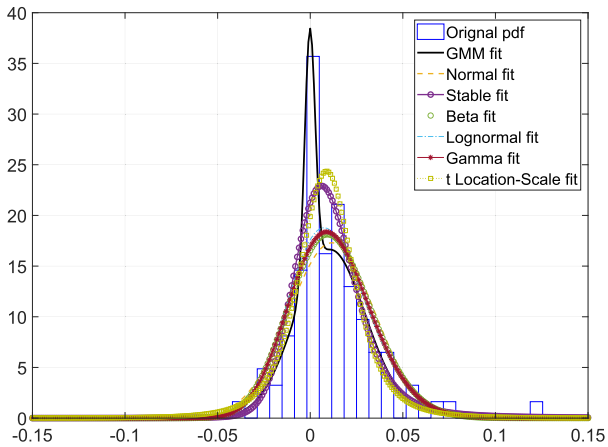


Fig. 8. Comparison of the fitting performance between GMM and other distributions applied to the wind updated forecasts for forecast period H12Q1 across all 365 paths observed at the last time step.

Table 7
The GMM offers a better fit for the density histogram as compared to other distributions based on the χ^2-statistics.

Distribution	GMM	Normal	Stable	Beta	Log-normal	Gamma	t-Location-Scale
Value	788.21	2596.51	845.59	10651.66	15651.71	11971.54	6263.29

returns more efficiently, by removing zero returns. In our case, p = (Total number of returns - The number of zero returns)/Total number of returns. The GMM distributions of all time steps of wind H12Q1 are shown in Fig. 9. We observe that the kurtosis of the distribution is high at the beginning of the evolution, while the kurtosis is lower before the delivery time, as shown in Section 2. In other words, the infeed forecasts become more volatile during the last 3 h before the delivery time. Besides, outliers appear as individual components, as seen in Fig. 9. This reflects the GMM's flexibility of fitting very differently shaped distributions.

4.2.2. Scenario generation

Usually, a joint distribution of random variables cannot guarantee the independence of the multiple component distributions. This is the reason why it is difficult to sample from a joint distribution. The same issue also exists in GMMs. The idea to circumvent this obstacle is to sample individually from each Gaussian distribution components, and then collect all samples to generate a joint distribution. The procedure of generating scenarios from a GMM is displayed in the following three steps:

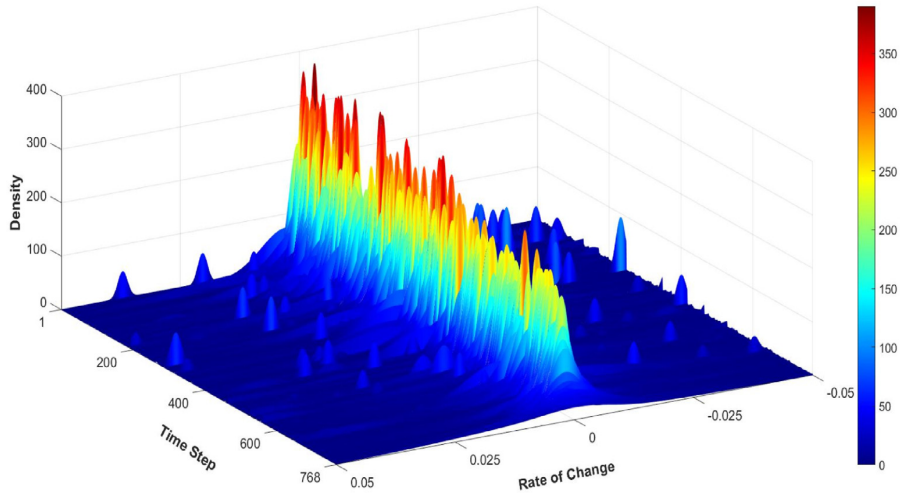


Fig. 9. The GMM fit for quarter-hourly updated wind forecasts for forecast period H12Q1 across all 365 paths at all time steps.

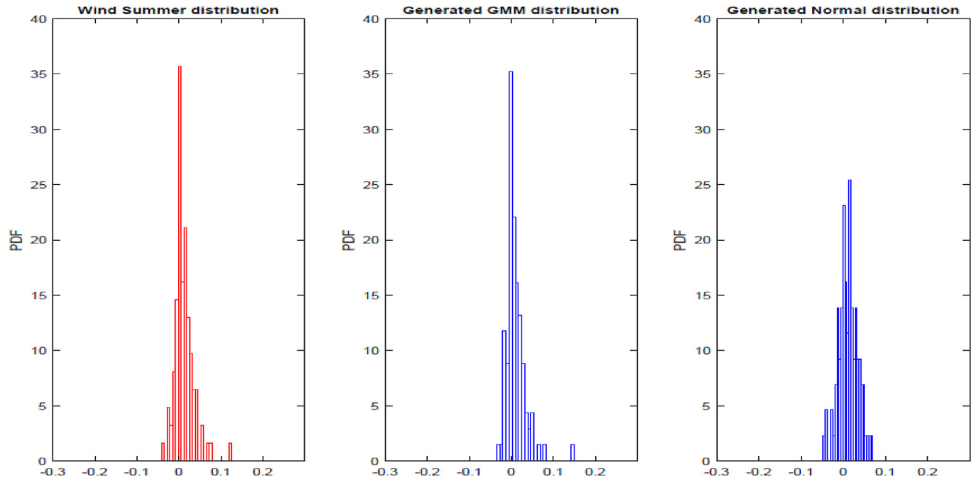


Fig. 10. The histograms of returns observed at the last time step for forecast period H12Q1, wind: the empirical distribution, GMM generated scenarios and Normal distribution generated scenarios.

1. Based on the weights of components of a GMM, the number of samples generated from each component is assigned.
2. Sampling from multiple Gaussian distribution components.
3. Assembling all those samples to obtain a new sample set.

Sampling from the individual series of Gaussian distributions is the only requirement to generate GMM scenarios. The MATLAB also provides a sampling function for GMM, which can generate samples within milliseconds. In Fig. 10, we show comparatively the histograms of observed wind infeed in summer for H12Q1 at the last forecasting time step versus the generated scenarios with GMM and a normal distribution, respectively. We observe a clear success of GMM in fitting the original distribution.

5. Simulations

In this section, we show a simulation exercise to describe the evolution of renewables infeed updated forecasts across the 365 paths. In particular, we test the modeling performance by applying a cross-validation procedure to the paths of wind forecasts for H12Q1 as forecast period, which is observed each day in 2015. The basic idea of cross-validation is to repeatedly train the models on a major part of the data and evaluate them on the remaining part. The repeated evaluation for different subsets will make the

Table 8
The χ^2 statistic of the GMM, normal, stable, lognormal and gamma distributions for the parameters of stochastic models. The GMM provides a better fit for the density histogram as compared to the other distributions.

	GMM	Normal	Stable	Log-normal	Gamma	t-Location-Scale
θ_1^{OU}	0.81	37.84	13.68	32.41	5.27	26.38
θ_2^{OU}	2.47	109.55	15.63	783.58	259.46	58.99
θ_3^{OU}	1.73	27.71	7.78	6.87	13.58	28.85
θ_1^{CKLS}	1.34	34.08	19.62	12.62	3.65	43.94
θ_2^{CKLS}	1.33	76.25	9.16	789.64	249.08	47.18
θ_3^{CKLS}	1.38	28.96	7.06	4.61	7.99	18.31
θ_4^{CKLS}	2.18	18.91	8.15	676.11	153.53	9.80
θ_1^{CIR}	1.76	32.79	15.15	14.73	1.74	29.10
θ_2^{CIR}	1.37	65.50	3.27	728.23	239.24	35.26
θ_3^{CIR}	1.02	15.88	3.06	3.93	4.26	2.83

results less variable to estimate the skill of models even with a small dataset (Messner et al., 2020).

5.1. Stochastic models

We propose two strategies for the simulations exercise when applying the stochastic models. The first strategy follows a standard stochastic simulation process. We calibrate each stochastic model to each one of the 364 paths and generate 1000 scenarios for each path (total 364,000 scenarios). Thus, we utilize all forecast information from the 364 paths. Then, we cluster 1000 paths from generated scenarios based on the closeness for the remaining path (the 365th path). We generate 1000 scenarios as in the study of Narajewski and Ziel (2020b), in which 1000 forecasts were generated for ensemble forecasting evaluation. Besides, distance-based scenarios reduction is a common technique in stochastic programming (Horejšová et al., 2020; Dvorkin et al., 2014). For instance, k-means clustering is one commonly used scenario reduction method. In our case, we argue that the clustered 1000 scenarios are the most similar paths to the 365th path, which are used to assess the simulation performance of the proposed model by employing multiple measure metrics. We repeat this evaluation process 365 times (from the 1st path to the 365th path).

Next, based on the parameter distributions across all 365 paths as shown in Section 4, Fig. 6, we propose the second simulation strategy, the parameter sampling method, to generate scenarios. Before conducting a simulation, we employ several probability models to fit the parameter distributions and look for the best fit model. The test results for the χ^2 goodness-of-fit statistic of the GMM-, normal-, stable-, lognormal-, gamma- and t-Location-Scale distributions are displayed in Table 8, and the parameters θ_1, θ_2 and θ_3 of the OU model are illustrated in Fig. 11.[5] GMM fits best the parameter distributions. It is worthy to notice the gamma-, stable- and t-Location-Scale distributions perform well, which is not surprising, since they are commonly employed to model weather forecasts in prior studies (Tewari et al., 2011; Bruninx and Delarue, 2014; Menemenlis et al., 2012; Peña et al., 2020). However, given the versatile distributions of parameters, GMM is a better choice. For the cross-validation, we fit GMM across 364 paths to each parameter of the stochastic models, and then, based on the fitted GMM, sample a value for each parameter and build a synthesized stochastic differential equation for each stochastic model. Each equation is used to generate a scenario. We repeatedly synthesize 364,000 sets of parameters and build 364,000 equations, and correspondingly generate 364,000 scenarios for each model. Then, we cluster 1000 closest scenarios for the 365th path to assess the simulation performance for each model. This process is replicated 365 times.

5.2. GMM models

We apply the GMM method (probabilistic model) to simulate the renewables infeed forecasts as introduced in Sections 3 and 4. In particular, for each time step t of a specific path, we sample a return of the updated forecasts, r_t, from the returns' distribution simulated by GMM based on the information from the other 364 paths. Then, based on the simulated wind forecast at time t, y_t, and r_t, we obtain $y_{t+1} = y_t \times (1 + r_t)$. For each path, we generate 364,000 scenarios and then select 1000 closest ones to conduct validation.

5.3. Simple benchmark models

We also apply a random walk (RW) model with drift (stochastic model) as the first benchmark model. The parameters of the RW model are calibrated by Autoregressive Integrated Moving Average Model (ARIMA(0,1,0)). For the RW model, the simulation evaluation procedure is analogous to the stochastic models introduced in Section 5.1. We don't consider parameter sampling for the RW.

As a second benchmark model, we employ a Naive approach by using all observations of updated renewables forecasts from the

[5] The illustration of the parameters distributions of CKLS and CIR models can be found in Fig. A.17 and A.18.

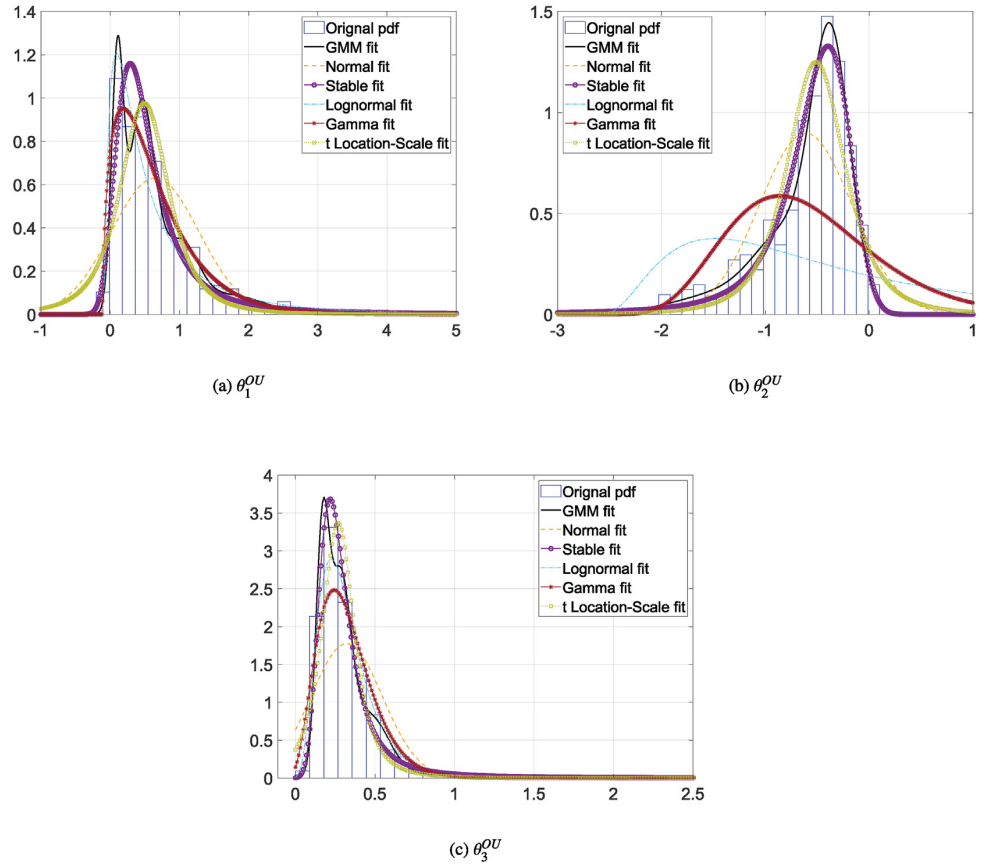


Fig. 11. GMM and other distributions fit for the parameters' distributions of the OU model.

first step to the last to forecast the next value. In particular, for validation, we generate 1000 scenarios for each of the 364 paths by randomly sampling from the values of the true trajectories.

5.4. Simulation performance

In order to evaluate the effectiveness of the generated simulations, we employ four measure metrics. The first one is the mean absolute error (MAE). It is a strictly proper scoring rule for median evaluation. In our case study, the MAE is defined as follows (Narajewski and Ziel, 2020b):

$$\text{MAE} = \frac{1}{NT}\sum_{d=1}^{N}\sum_{t=1}^{T}\left|median_{m=1,\dots,M}(\hat{y}_{t,m}^{d}) - y_t^d\right| \tag{16}$$

where $N = 365$, $T = 768$, $M = 1000$, y_t^d is the representation of samples, $\hat{y}_{t,m}^d$ is the m-th simulations of y_t^d and $median_{m=1,\dots,M}(\hat{y}_{t,m}^d)$ is the median of M simulations.

The second metric is the root mean absolute error (RMSE). The main difference between MAE and RMSE is that the RMSE gives a relatively high weight to large errors, since errors are squared before they are averaged. The RMSE is the strictly proper scoring rule for mean evaluation. The RMSE is defined as follows (Narajewski and Ziel, 2020b):

$$\text{RMSE} = \sqrt{\frac{1}{NT}\sum_{d=1}^{N}\sum_{t=1}^{T}\left(\frac{1}{M}\sum_{m=1}^{M}\hat{y}_{t,m}^{d} - y_t^d\right)} \tag{17}$$

where $N = 365$, $T = 768$, $M = 1000$, y_t^d is the representation of samples and $\hat{y}_{t,m}^d$ is the m-th simulations of y_t^d.

Table 9
MAE, RMSE and CRPS measures of the considered models for simulations. ∗(P) stands for parameter sampling method.

Model	OU	CKLS	CIR	OU(P)	CKLS(P)	CIR(P)	GMM	RW	Naive
MAE	0.1392	0.1372	0.1346	0.1413	0.1375	0.1327	**0.1047**	0.1321	0.2186
RMSE	0.2760	0.2722	0.2719	0.2623	0.2438	0.2444	**0.2032**	0.2568	0.4486
CRPS	0.1006	0.0980	0.0969	0.1032	0.0980	0.0959	**0.0785**	0.0966	0.1705

The continuous ranked probability score (CRPS), the third metric, is one of the most common single-valued scores to evaluate the marginal density fit of continuous scenarios (Gneiting and Raftery, 2007; Narajewski and Ziel, 2020b; Nowotarski and Weron, 2018). A formula for forecasts given as ensembles can be written as (Messner et al., 2020):

$$\text{CRPS} = \frac{1}{NT}\sum_{d=1}^{N}\sum_{t=1}^{T}\left(\frac{1}{M}\sum_{m=1}^{M}|\hat{y}_{t,m}^{d} - y_{t}^{d}| - \frac{1}{2M^{2}}\sum_{m=1}^{M}\sum_{l=1}^{M}|\hat{y}_{t,m}^{d} - \hat{y}_{t,l}^{d}|\right) \tag{18}$$

where $N = 365$, $T = 768$, $M = 1000$, y_t^d is the representation of samples and $\hat{y}_{t,m}^{d}$, $\hat{y}_{t,l}^{d}$ are the m-th, the l-th independent simulations of y_t^d, respectively.

The fourth one is the band depth or the modified band depth rank method introduced by Thorarinsdottir et al. (2016). The concept of band depth is derived from López-Pintado and Romo (2009). In general, it allows for ranking a sample of data from the center outwards and defines the centrality or outlyingness of an observation. If the simulations or forecasts lack calibration, the shape of the verification rank histogram reveals the nature of the misspecification. For instance, a ∪-shaped histogram is an indication of a lack of correlation in the simulations or forecasts, while a hump or ∩-shape suggests too high correlation. A skewed histogram with too many high ranks indicates an overdispersive simulations- or forecasts ensemble, while too many low ranks indicate an underdispersive or biased ensemble (Thorarinsdottir et al., 2016). Essentially, the observations should be indistinguishable from random draws from the generated distributions if the model is well-calibrated. The ideal model is the one with a rank that has a uniform distribution. In this paper, we applied the band depth method to critically assess the calibration of different models.

The value of applied error measures is shown in Table 9. Among all methods, the performance of GMM is the best. Results imply that there is no significant difference in performance among the proposed stochastic models. Also, the stochastic models do not perform better than the benchmark stochastic RW with a drift model. This indicates that generality of the scenarios is essential for simulation. Results also denote that the Naive methods are not suitable for simulation in this case. Similar evidence can be perceived in the simulation performance of the proposed models for PV H12Q1 infeed forecasts, which is shown in Appendix: Table A.12.

In addition, we applied the band depth method to quantify the simulation performance of different models. We define $\mathbf{S} = \{x_1, \ldots, x_{365}\}$ to denote a set of points in $\mathbb{R}^d$. This is to say that we assume $\mathbf{S}$ contains 365 elements (paths), of which $\{x_1, \ldots, x_{364}\}$ are the simulated ensemble members and the corresponding observation is the x_{365} path. Hence, in this case study, the number of bins is 365, d is 767, which is the number of time steps to generate simulations, and the results are based on 1000 repetitions for 365 observation (paths), respectively. In particular, for each observation, three stochastic models and two benchmark models generate each ensemble element based on each of the remaining 364 paths, respectively, while the stochastic parameter sampling method and the GMM generate 364 elements based on parameters' distribution and returns' distribution across the 364 paths, respectively. The results of simulation band depth tests for GMM, CIR, CKLS(P), and RW are displayed in Fig. 12.

From Fig. 12(a), as expected, we can see that the rank histogram of GMM is close to being uniform, with a slightly large region of "high ranks". The reason for the high ranks is that there are several observations (paths) as outliers (experience severe fluctuations during some periods or a significant correction from the first forecast), deviating from the other observations, as shown in Fig. 2. Simulations generated based on the information from other observations (not outliers) will be overdispersive for these outliers. The uniformly-distributed rank histogram of GMM indicates that there are no obvious effects of bias or under/over-dispersion. However, the rank histograms of the CIR model show a strong ∩-shape, as shown in Fig. 12(b). The ∩-shape histogram indicates that there are less inlying and less outlying observations than would be expected from a sample. Meanwhile, the rank histograms of the CKLS(P) and RW model have triangular shapes with too many high/low ranks, as shown in Fig. 12(c) and (d), indicating that they have an overdispersive/underdispersive simulation ensemble.

The somewhat uniformity of the verification rank histogram is a necessary condition to assess the models'simulation performance. Among all the calibration metrics, the histogram of the GMM generated observations appears close to a uniform distribution, which indicates a high performance of the method to simulate the series of renewables updated forecasts. The result of the band depth is reciprocally confirmed when employing the MSE, RMSE, and CRPS, which suggests that the GMM method is the best choice for simulations of wind and photovoltaic infeed forecasts, outperforming the other proposed models.

6. Out-of-sample performance

In this section, we test the ensemble prediction performance of all models proposed above on the evolution of wind updated forecasts for H12Q1, following an expanding window procedure. The initial in-sample size comprises the first 672 time steps, while the out-of-sample the remaining 96 quarter-hourly updated forecasts, which are highly relevant for the intraday trading. As shown in Fig. 1, the trading day starts at the step 673. Also, as shown in Figs. 2(a) and 3(c), we observe that weather updates are more frequent closer to the forecast period, enhancing the need of traders to balance out (updated) forecasting errors. Furthermore, the study of

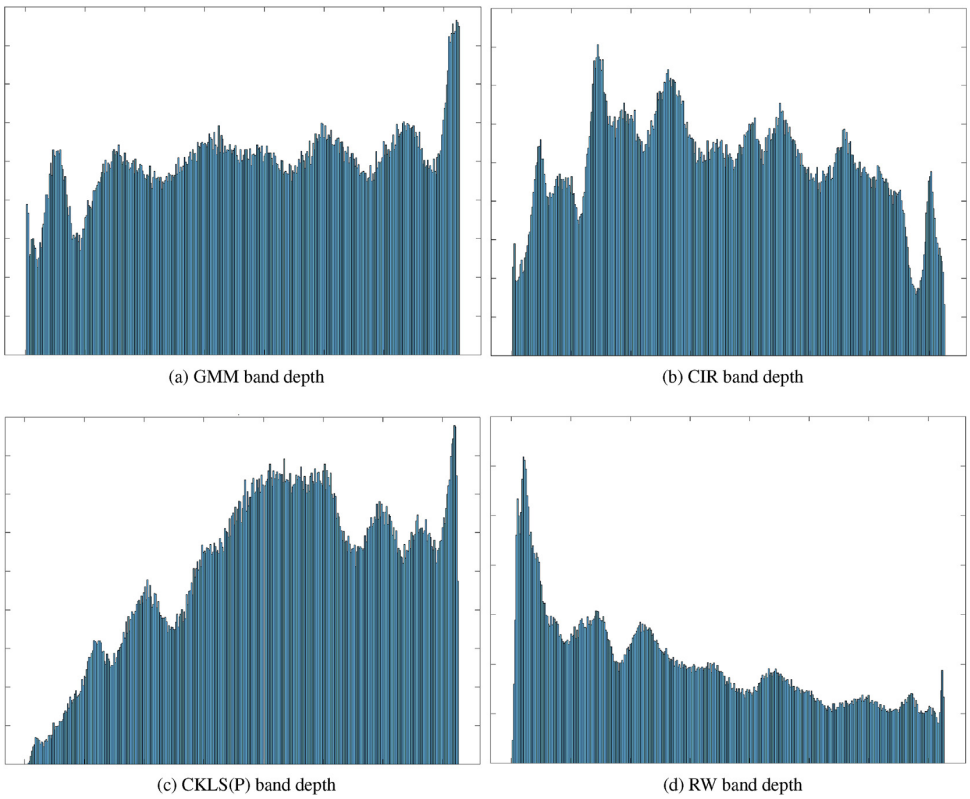


Fig. 12. Simulation band depth tests of the wind updated forecasts for H12Q1 as forecast period, for GMM, CIR, CKLS(p), and RW models, respectively. ∗(P) stands for parameter sampling method.

Narajewski and Ziel (2020b) shows that the closer to the delivery period, the greater the transaction frequency. The forecasting exercise is done for each one of the 365 paths ($N = 365$) separately.

To evaluate the prediction performance of the proposed models, we employ a rolling window forecasting approach. In particular, our rolling window is defined as "expanding window", which has a fixed starting point and incorporates new data as it becomes available. In particular, for each path, we calibrate stochastic models using the information from the step 1 to the step 672 and make the first prediction for the step 673. For this purpose, we first generate 1000 values for step 673 for each stochastic model. Thus, at this time, we forecast 1000 values ($M = 1000$) for the infeed forecasts one step ahead. Then, we increase the length of the forecast trajectory by the generated renewables forecast at step 673 and based on this predict the value for the next step 674, continuing this recursive process until the delivery time. This process generates 1000 trajectories, each including 96 points ($T = 96$). Then, we move the window forward by one step, re-calibrate models by including the newly available information, and repeat the exercise until the end of out-of-sample data. A similar procedure is applied in the study of Narajewski and Ziel (2020b).

The probabilistic model GMM has a different mechanism to make a prediction compared to the stochastic models discussed in the paper. The GMM considers the information simultaneously across paths for given observation time steps. The essential assumption of GMM for forecasting is that, at a particular time step, the infeed forecasts update following a statistical pattern. Thus, the historical information across paths can be used to give an indication of the range of possible future forecasts. A similar application can be found in the study of Zhiwen et al. (2018), which employs GMM to model the wind power forecast errors. However, only one year of data is available in our case study. To deal with the issue of data limitation when applying GMM to predict an update in wind/PV forecast of a specific path at a given time step, we take the other 364 paths as the pseudo-historical data and fit GMM to those. Then, we take the fitted GMM to conduct ensemble forecasting as introduced above. The rolling window procedure of GMM is similar to that of the stochastic models. The value of utilized error measures is shown in Table 10.

From Table 10, we can observe that GMM performs better than the stochastic models. However, RW performs worse than the stochastic models. Thus, results indicate that the evolution of infeed forecasts follows a stochastic rather than a pure random walk process. The results are consistent with the observations and stationary tests. The poor performance of the Naive model indicates that only randomly repeating historical information is not efficient for forecasting.

Table 10
MAE, RMSE and CRPS measures of the considered models for the last day's (last 96 time step) ensemble prediction.

Model	OU	CKLS	CIR	GMM	RW	Navie
MAE	0.0669	0.0597	0.0618	**0.0568**	0.0667	0.2604
RMSE	0.1087	0.1053	0.1096	**0.1030**	0.1136	0.4399
CRPS	0.0522	0.0462	0.0503	**0.0439**	0.0542	0.1840

Table 11
MAE, RMSE and CRPS measures of the considered models for the last 3 h (12 time steps) ensemble prediction after re-modeling.

Model	OU	CKLS	CIR	GMM	RW	Naive
MAE	**0.0330**	0.0350	0.0331	0.0356	0.0347	0.0606
RMSE	**0.0603**	0.0858	**0.0603**	0.0630	0.0631	0.0925
CRPS	**0.0252**	0.0266	**0.0252**	0.0269	0.0263	0.0444

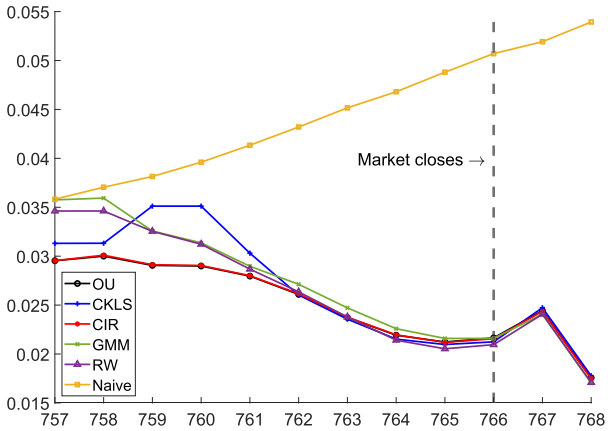


Fig. 13. CRPS measures of the considered models from the last 3 h (12 time steps) prediction after re-modeling.

As we observe from Figs. 2(a) and 3(c), the updates become frequent in the last 8.5 h. Thus, two dynamics exist in our forecasting performance which can be separated into two periods: low-frequency and high-frequency periods. To eliminate the influence of the low-frequency-update period on the stochastic models, we reestimate results for short-period forecasting. We only consider the information across 8.5 h before the delivery time (high-frequency-update period). In particular, we calibrate stochastic models to the data from the step 735 to 756 (22 time steps) and approach rolling window to forecast the last 12 time steps (3 h before the delivery and 2.5 h before the market closes, when updates in the wind forecasts are much more frequent). The test results of applied error measures are shown in Table 11 and the CRSP values at each time step in the out-of-sample for each model are shown in Fig. 13.

Based on the results in Table 11, we observe that the OU and CIR models perform better than the other models. This shows that the elimination of the information along the paths from the low-frequency-update period leads to a more accurate forecasting performance of stochastic models. Thus, as previously discussed, results indicate that in the period of frequent weather updates, the evolution of wind forecasts can be successfully modeled by the dynamics of stochastic processes. As shown in Fig. 13, the forecasting performance of stochastic models increases closer to the forecast period, which is intuitive, given their short-term autoregressive nature. We furthermore observe that the closer we are to the target delivery period, in contrast to the stochastic models, the Naive model version loses its forecast accuracy. It is not surprising to find that the more historical information included, the worse the prediction power of the Naive model. The result is in alignment with the observations that true trajectories are drifting. In addition, the proposed models' performance tends to converge before the delivery, as calibration window size increases and prediction window (forecast horizon) decreases. Equivalent evidence can be observed in the prediction performance of different models for PV H12Q1 infeed forecasts, which is shown in Appendix: Fig. A.19.

7. Conclusions

The integration of European electricity intraday markets triggers advanced research for improved trading strategies. Updated information of wind and photovoltaic infeed forecasts are essential supply-side variables for the adjustment positions in the intraday

trading. As shown in the study of Kremer et al. (2020a), renewable forecast updates are reflected in the prices of 15-min contracts within one trading minute. Furthermore, Kiesel and Paraschiv (2017) show that intraday prices adjust asymmetrically to forecasting errors in renewables. In particular, there is a 15-min periodicity in the liquidity pattern in intraday trading which originates from newly arriving updated renewables forecasts. Renewable forecasts are updated in 15-min intervals and traders continuously adjust their bids to updated information from the weather data suppliers. It becomes therefore increasingly significant to develop reliable methods to simulate and predict the evolution of wind and PV updates that are practically useful for electric power industries or for designing optimal trading strategies.

To the best of our knowledge, this is the first study that explores accurate forecasting and simulation models for the evolution of wind and PV updated forecasting errors, which are a useful tool to participants in the intraday market for building optimal trading strategies. Clearly, our results break ground for several applications to intraday trading. In particular, best simulators can support investors to test virtually any trading strategy at low risk. Besides, the accurate prediction of renewables forecasting errors can help power producers to better organize and adjust their production schedule, further decreasing the risk of blackout and voltage collapse.

In this paper, we tested the in- and out-of-sample fit of six several models applied to 365 paths of updated renewable forecasts, wind and PV, for a given quarter-hourly product, available in 15-min updates over eight days. In particular, we assess comparatively the performance of stochastic models chosen in line with the statistical patterns in the data versus a GMM model. Robustness checks suggest that GMM is a reliable tool to simulate the updates in the weather data, showing a superior simulation performance versus the classical stochastic models OU, CKLS and CIR. For the out-of-sample analysis, we find that the proposed models have different prediction performances, depending whether the weather forecast updates follow a low- or a high-frequency pattern. The latter is recognizable the closer we come to the delivery period. In particular, the GMM performs better than the stochastic models during the low-frequency-update period. However, in the high-frequency-update period, when time approaches the forecast period, stochastic models show superior performance, as updated renewables forecasts show an accentuated autoregressive nature. We furthermore show the distribution of model parameters over the 365 analyzed paths of renewables forecasts, which can be explored in stress-testing exercises or in hedging upside/downside risk. In further research, one can explore this prior information of parameters in a Bayesian approach. Furthermore, accurate models for wind and PV updated forecasting errors can be used to enhance existing econometric models for intraday electricity prices. Simulations performed in this study are useful input to stochastic programming applications for optimal electricity production planning.

Credit author statement

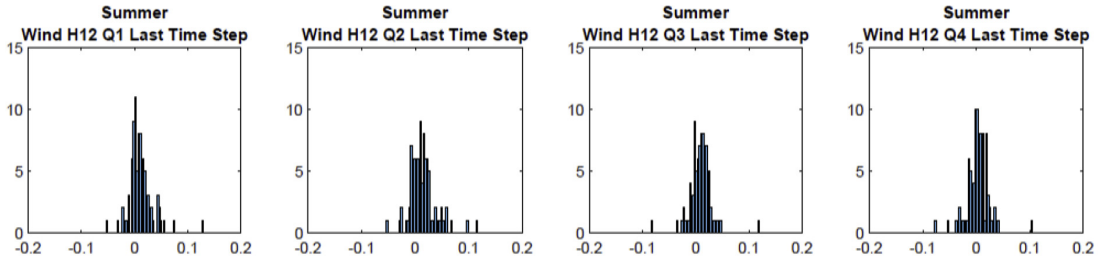
Wei Li: Conceptualization, Data curation, Formal analysis, Methodology, Software, Resources. Florentina Paraschiv: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Validation.

Acknowledgment

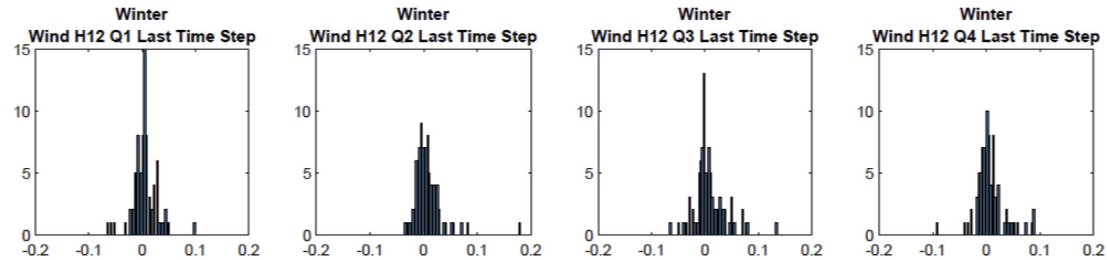
This work acknowledges research support by COST Action “Fintech and Artificial Intelligence in Finance - Towards a transparent financial industry” (FinAI) CA19130. This research has been performed within the +CityxChange[6] (Positive City ExChange) project under the Smart Cities and Communities topic that has received funding from the European Union’s Horizon 2020 research and innovation programme under Grant Agreement No. 824260. Florentina Paraschiv thanks the funding from *Adolf Øiens Donasjonsfond* Energizing New Computational Frontiers and the Isaac Newton Institute for Mathematical Sciences for its hospitality during the programme “The mathematics of energy systems” which was supported by EPSRC Grant Number EP/R014604/1. Critical comments and advice from Fred Espen Benth, Michael Coulon, Stein-Erik Fleten, Pierre Pinson and Ranik Raaen Wahlstrøm are gratefully acknowledged. Special thanks to Ruediger Kiesel and EWE TRADING GmbH [7] for providing the data for the research analysis. The renewable power infeed forecast data are obtained via a (Uni-)Vendor contract and cannot be made publicly available due to contractual conditions.

[6] https://cityxchange.eu/.
[7] https://www.ewe.com/en.

Appendix A

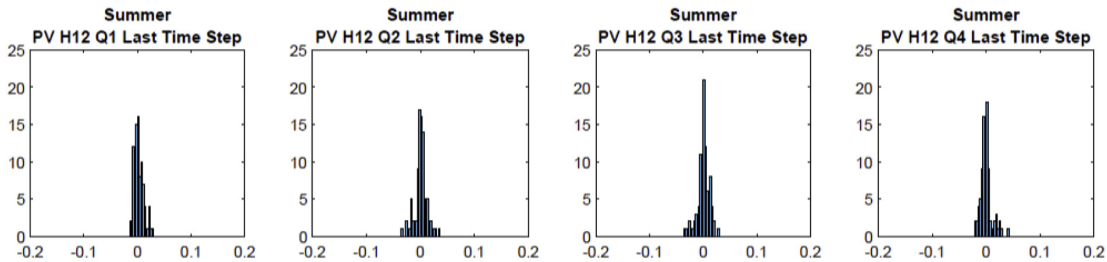


(a) Summer Wind Last Time Step

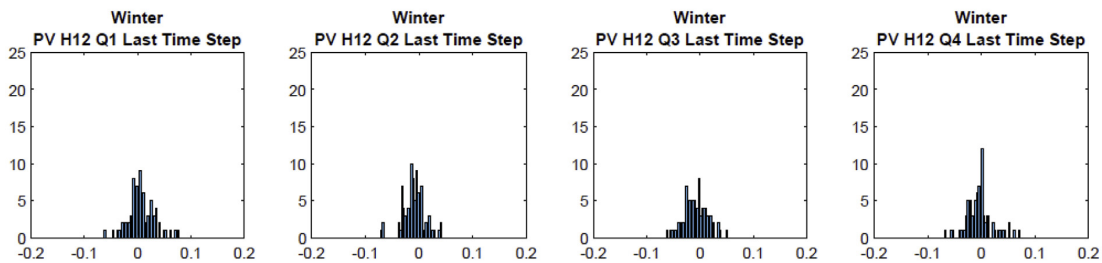


(b) Winter Wind Last Time Step

Fig. A.14 The distribution of the last time step of Hour 12 from Quarter 1 to 4 of the summer and winter wind infeed forecasts in 2015.



(a) Summer PV Last Time Step



(b) Winter PV Last Time Step

Fig. A.15 The distribution of returns for the last time step for Hour 12 from Quarter 1 to 4 of the summer and winter PV infeed forecasts observed in 2015.

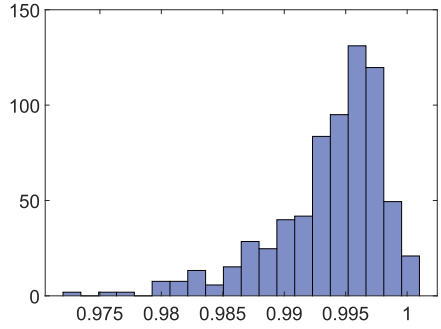
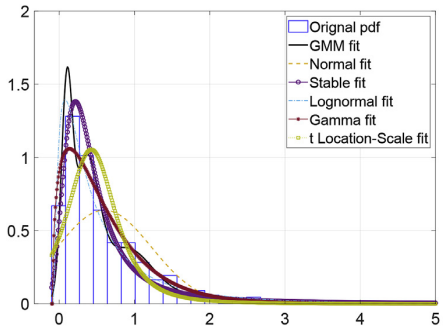
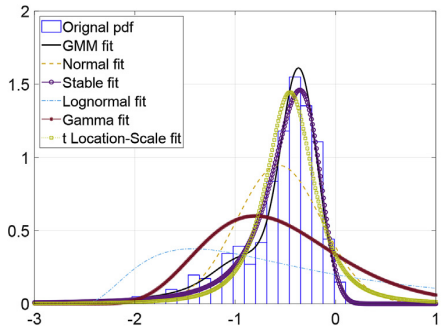


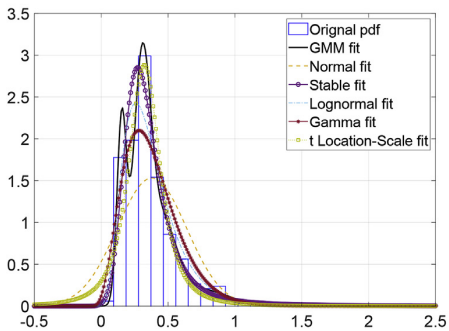
Fig. A.16 The distribution of φ parameters across all 365 paths of wind updated forecasts.



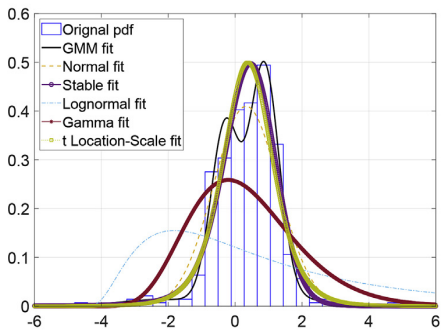
(a) θ_1^{CKLS}



(b) θ_2^{CKLS}

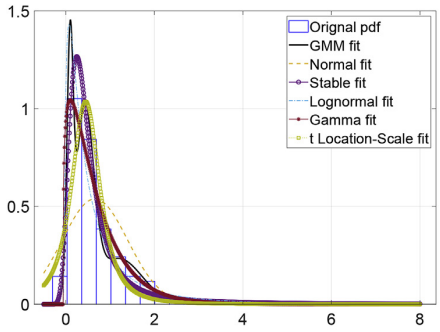


(c) θ_3^{CKLS}

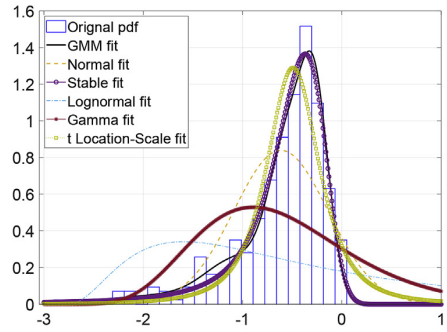


(d) θ_4^{CKLS}

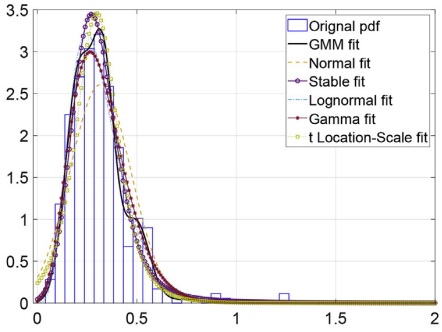
Fig. A.17 Parameter distributions of CKLS model.



(a) θ_1^{CIR}



(b) $\theta_2^{C}IR$



(c) θ_3^{CIR}

Fig. A.18 Parameter distributions of CIR model.

Table A.12
MAE, RMSE and CRPS measures of the considered models for simulations (PV H12Q1). ∗(P) stands for parameter sampling method.

Model	OU	CKLS	CIR	OU(P)	CKLS(P)	CIR(P)	GMM	RW	Naive
MAE	0.1279	0.1276	0.1286	0.1303	0.1335	0.1295	**0.0962**	0.1196	0.1791
RMSE	0.2641	0.2540	0.2527	0.2462	0.2432	0.2397	**0.1864**	0.2354	0.3618
CRPS	0.0946	0.0937	0.0942	0.0977	0.1000	0.0960	**0.0699**	0.0906	0.1414

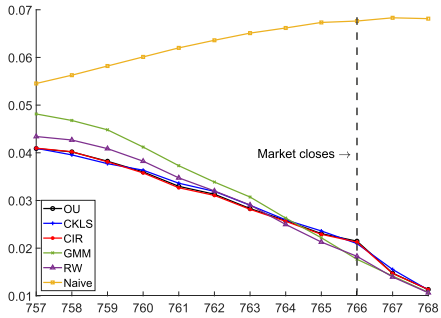


Fig. A.19 CRPS measures of the considered models from the last 3 h (12 time steps) prediction after re-modeling (PV H12Q1).

References

Alain, B., Alexandre, B., 2015. Cox–ingersoll–ross model for wind speed modeling and forecasting. Wind Energy 19, 1355–1365, https://doi.org/10.1002/we.1896.

Benth, F., Ibrahim, N., 2017. Stochastic modeling of photovoltaic power generation and electricity price. J. Energy Mark. 10, 1–33, https://doi.org/10.21314/JEM.2017.164.

Benth, F., Kallsen, J., Meyer-Brandis, T., 2007. A non-Gaussian ornstein-uhlenbeck process for electricity spot price modeling and derivatives pricing. Appl. Math. Finance 14, 153–169, https://doi.org/10.1080/13504860600725031.

Benth, F., Pircalabu, A., 2018. A non-Gaussian ornsteinuhlenbeck model for pricing wind power futures. Appl. Math. Finance 25, 36–65, https://doi.org/10.1080/1350486X.2018.1438904.

Bibbona, E., Panfilo, G., Tavella, P., 2008. The ornstein–uhlenbeck process as a model of a low pass filtered white noise. Metrologia 45, 117–126, https://doi.org/10.1088/0026-1394/45/6/s17.

Boland, J., 2008. Time series modelling of solar radiation. In: Badescu, V. (Ed.), Modeling Solar Radiation at the Earth's Surface: Recent Advances. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 283–312, https://doi.org/10.1007/978-3-540-77455-6_11.

Bruninx, K., Delarue, E., 2014. A statistical description of the error on wind power forecasts for probabilistic reserve sizing. IEEE Trans. Sustain. Energy 5, 995–1002, https://doi.org/10.1109/TSTE.2014.2320193.

Dvorkin, Y., Wang, Y., Pandzic, H., Kirschen, D., 2014. Comparison of scenario reduction techniques for the stochastic unit commitment. In: 2014 IEEE PES General Meeting | Conference Exposition, pp. 1–5, https://doi.org/10.1109/PESGM.2014.6939042.

Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50, 987–1007.

Epexspot, 2019. Market Coupling Xbid: Cross-Border Intraday Market Project. https://www.epexspot.com/en/market-coupling/xbid_cross_border_intraday_market_project. (Accessed 26 November 2019).

Frauendorfer, K., Paraschiv, F., Schürle, M., 2018. Cross-border effects on swiss electricity prices in the light of the energy transition. Energies 11, https://doi.org/10.3390/en11092188.

Gibson, R., Schwartz, S.,E., 1990. A non-Gaussian ornstein-uhlenbeck process for electricity spot price modeling and derivatives pricing. J. Finance 45, 959–976, https://doi.org/10.1111/j.1540-6261.1990.tb05114.x.

Glasserman, P., 2004. Monte Carlo Methods in Financial Engineering, vol. 53. Springer, New York, https://doi.org/10.1007/978-0-387-21617-1.

Gneiting, T., Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. J. Am. Stat. Assoc. 102, 359–378, https://doi.org/10.1198/016214506000001437.

Guidoum, C.,A., Boukhetala, K., 2019. Parametric Estimation of 1-d Stochastic Differential Equation. https://cran.r-project.org/web/packages/Sim.DiffProc/vignettes/fitsde.html. (Accessed 26 November 2019).

Hagfors, L.I., Kamperud, H.H., Paraschiv, F., Prokopczuk, M., Sator, A., Westgaard, S., 2016a. Prediction of extreme price occurrences in the German day-ahead electricity market. Quant. Finance 16, 1929–1948, https://doi.org/10.1080/14697688. 2016.1211794.

Hagfors, Lars Ivar, Paraschiv, Florentina, Molnar, Peter, Westgaard, Sjur, 2016b. Using quantile regression to analyze the effect of renewables on eex price formation. Renew. Energy Environ. Sustain. 1, 32, https://doi.org/10.1051/rees/2016036.

Horejšová, M., Vitali, S., Kopa, M., Moriggia, V., 2020. Evaluation of scenario reduction algorithms with nested distance. Comput. Manag. Sci. 17, 241–275, https://doi.org/10.1007/s10287-020-00375-4.

Janczura, J., Orze, S., Wylomanska, A., 2011. Subordinated alpha-stable ornstein-uhlenbeck process as a tool for financial data description. Physica A. 390, 4379–4387. https://econpapers.repec.org/paper/wuuwpaper/hsc1103.htm.

Janke, T., Steinke, F., 2019. Forecasting the price distribution of continuous intraday electricity trading. Energies 12, https://doi.org/10.3390/en12224262.

Kath, C., Ziel, F., 2018. The value of forecasts: quantifying the economic gains of accurate quarter-hourly electricity price forecasts. Energy Econ. 76, 411–423, https://doi.org/10.1016/j.eneco.2018.10.005.

Kiesel, R., Paraschiv, F., 2017. Econometric analysis of 15-minute intraday electricity prices. Energy Econ. 64, 77–90, https://doi.org/10.1016/j.eneco.2017.03.002.

Kremer, M., Kiesel, R., Paraschiv, F., 2020a. A Fundamental Model for Intraday Electricity Trading. Philosophical Transactions of the Royal Society A, https://doi.org/10.2139/ssrn.3489214. Forthcoming.

Kremer, M., Kiesel, R., Paraschiv, F., 2020b. Intraday electricity pricing of night contracts. Energies 13, https://doi.org/10.3390/en13174501.

Liu, B., Nowotarski, J., Hong, T., Weron, R., 2017. Probabilistic load forecasting via quantile regression averaging on sister forecasts. IEEE Trans. Smart Grid 8, 730–737, https://doi.org/10.1109/TSG.2015.2437877.

López-Pintado, S., Romo, J., 2009. On the concept of depth for functional data. J. Am. Stat. Assoc. 104, 718–734, https://doi.org/10.1198/jasa.2009.0108.

Maciejowska, K., Nitka, W., Weron, T., 2019. Day-ahead vs. intraday—forecasting the price spread to maximize economic benefits. Energies. 12, https://www.mdpi.com/1996-1073/12/4/631. 10.3390/en12040631.

Maciejowska, K., Nowotarski, J., Weron, R., 2016. Probabilistic forecasting of electricity spot prices using factor quantile regression averaging. Int. J. Forecast. 32, 957–965, https://doi.org/10.1016/j.ijforecast.2014.12.004.

Menemenlis, N., Huneault, M., Robitaille, A., 2012. Computation of dynamic operating balancing reserve for wind power integration for the time-horizon 1–48 hours. IEEE Trans. Sustain. Energy 3, 692–702, https://doi.org/10.1109/TSTE. 2011.2181878.

Messner, J.W., Pinson, P., Browell, J., Bjerregård, M.B., Schicker, I., 2020. Evaluation of wind power forecasts—an up-to-date view. Wind Energy 23, 1461–1481, https://doi.org/10.1002/we.2497.

Monteiro, C., Ramirez-Rosado, I.J., Fernandez-Jimenez, L.A., Conde, P., 2016. Short-term price forecasting models based on artificial neural networks for intraday sessions in the iberian electricity market. Energies 9, https://doi.org/10.3390/en9090721.

Narajewski, M., Ziel, F., 2020a. Econometric modelling and forecasting of intraday electricity prices. J. Commod. Mark. 19, 100107, https://doi.org/10.1016/j.jcomm.2019.100107.

Narajewski, M., Ziel, F., 2020b. Ensemble forecasting for intraday electricity prices: simulating trajectories. Appl. Energy 279, 115801, https://doi.org/10.1016/j.apenergy.2020.115801.

Nowotarski, J., Weron, R., 2018. Recent advances in electricity price forecasting: a review of probabilistic forecasting. Renew. Sustain. Energy Rev. 81, 1548–1568, https://doi.org/10.1016/j.rser.2017.05.234.

Paraschiv, F., Bunn, D.W., Westgaard, S., 2016a. Estimation and Application of Fully Parametric Multifactor Quantile Regression with Dynamic Coefficients. University of St. Gallen, School of Finance Research. Paper No. 2016/07 https://ssrn.com/abstract= 2741692.

Paraschiv, F., Hadzi-Mishev, R., Keles, D., 2016b. Extreme value theory for heavy-tails in electricity prices. J. Energy Mark. 9, 21–50. https://ssrn.com/abstract=2795006.

Peña, J.I., Rodríguez, R., Mayoral, S., 2020. Tail risk of electricity futures. Energy Econ. 91, 104886, https://doi.org/10.1016/j.eneco.2020.104886.

Reynolds, D., 2009. Gaussian mixture models. In: Li, S.Z., Jain, A. (Eds.), Encyclopedia of Biometrics. Springer US, Boston, MA, pp. 659–663, https://doi.org/10.1007/978-0-387-73003-5_196.

Shoji, I., Ozaki, T., 1998. Estimation for nonlinear stochastic differential equations by a local linearization method. Stoch. Anal. Appl. 16, 733–752, https://doi.org/10.1080/07362999808809559.

Singh, R., Pal, C.,B., Jabr, A.,R., 2010. Statistical representation of distribution system loads using Gaussian mixture model. IEEE Trans. Power Syst. 25, 29–37, https://doi.org/10.1109/TPWRS.2009.2030271.

Tewari, S., Geyer, C.J., Mohan, N., 2011. A statistical model for wind power forecast error and its application to the estimation of penalties in liberalized markets. IEEE Trans. Power Syst. 26, 2031–2039, https://doi.org/10.1109/TPWRS.2011.2141159.

Thorarinsdottir, T.L., Scheuerer, M., Heinz, C., 2016. Assessing the calibration of high-dimensional ensemble forecasts using rank histograms. J. Comput. Graph Stat. 25, 105–122, https://doi.org/10.1080/10618600.2014.977447.

Uniejewski, B., Marcjasz, G., Weron, R., 2019. Understanding intraday electricity markets: variable selection and very short-term price forecasting using lasso. Int. J. Forecast. 35, 1533–1547, https://doi.org/10.1016/j.ijforecast.2019.02.001.

Valverde, G., Saric, T.,A., Terzija, V., 2012. Probabilistic load flow with non-Gaussian correlated random variables using Gaussian mixture models. IET Gener., Transm. Distrib. 6, 701–709, https://doi.org/10.1049/iet-gtd.2011.0545.

Weron, R., 2006. Modeling and Forecasting Electricity Loads and Prices: a Statistical Approach. Wiley, Chichester.

Weron, R., 2014. Electricity price forecasting: a review of the state-of-the-art with a look into the future. Int. J. Forecast. 30, 1030–1081, https://doi.org/10.1016/j.ijforecast.2014.08.008.

Zhiwen, W., Shen, C., Feng, L., 2018. A conditional model of wind power forecast errors and its application in scenario generation. Appl. Energy 212, 771–785, https://doi.org/10.1016/j.apenergy.2017.12.039.

Ziel, F., 2017. Modeling the impact of wind and solar power forecasting errors on intraday electricity prices. In: 2017 14th International Conference on the European Energy Market (EEM), pp. 1–5.

Zárate-Miñano, R., Anghel, M., Milano, F., 2013. Continuous wind speed models based on stochastic differential equations. Appl. Energy 104, 42–49, https://doi.org/10.1016/j.apenergy.2012.10.064.

Article B

Title:
Day-ahead electricity price prediction applying hybrid models of LSTM-based deep learning methods and feature selection algorithms under consideration of market coupling

Authors:
Wei Li
Denis Mike Becker

Reference:
Li W, Becker DM. Day-ahead electricity price prediction applying hybrid models of lstm-based deep learning methods and feature selection algorithms under consideration of market coupling. Energy: 121543, 2021. doi: https://doi:org/10:1016/j:energy:2021:121543.

Energy 237 (2021) 121543

Contents lists available at ScienceDirect

Energy

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

Day-ahead electricity price prediction applying hybrid models of LSTM-based deep learning methods and feature selection algorithms under consideration of market coupling

Wei Li*, Denis Mike Becker

NTNU Business School, Norwegian University of Science and Technology, 7491, Trondheim, Norway

ARTICLE INFO

Article history:
Received 7 December 2020
Received in revised form
13 July 2021
Accepted 17 July 2021
Available online 26 July 2021

Keywords:
Deep learning
Electricity price forecasting (EPF)
Electricity market coupling
Feature selection
Long short-term memory (LSTM)
The Nord Pool system price

ABSTRACT

The availability of accurate day-ahead electricity price forecasts is pivotal for electricity market participants. In the context of trade liberalisation and market harmonisation in the European markets, accurate price forecasting becomes difficult for electricity market participants to obtain because electricity forecasting requires the consideration of features from ever-growing coupling markets. This study provides a method of exploring the influence of market coupling on electricity price prediction. We apply state-of-the-art long short-term memory (LSTM) deep neural networks combined with feature selection algorithms for electricity price prediction under the consideration of market coupling. LSTM models have a good performance in handling nonlinear and complex problems and processing time series data. In our empirical study of the Nordic market, the proposed models obtain considerably accurate results. The results show that feature selection is essential to achieving accurate prediction, and features from integrated markets have an impact on prediction. The feature importance analysis implies that the German market has a salient role in the price generation of Nord Pool.

1. Introduction

Over the last two decades, worldwide energy markets have experienced a transition towards deregulation and harmonisation [1]. Under trade liberalisation, the traditional vertically integrated power utilities are replaced with decentralised business entities whose targets are to maximise their profits. Consequently, a growing number of market participants are exposed to intense competition, and their need for suitable decision support models to increase margins and reduce risk has significantly increased [2]. Thus, the availability of accurate day-ahead electricity price forecasts is vital for market participants to adjust production plans and to perform effective bidding strategies to make an economic profit. However, due to the productive structure and characteristics of electricity prices, highly accurate forecasting is quite challenging [3,4]. With the increasing integration of electricity markets, making accurate forecasts becomes even more difficult in the complex and integrated system. This is because the forecasting of electricity prices needs to consider a large number of factors from an ever-growing number of interconnected, neighbouring power markets. These factors include electricity prices, production, consumption, and other important features that influence cross-border electricity markets.

Numerous research efforts have contributed to the exploitation and development of advanced technologies for day-ahead electricity price forecasting (EPF), aimed at highly accurate forecasting results [5,6]. A considerable amount of literature has been devoted to EPF models, which can be classified into the following five categories [5]: multi-agent [7,8], fundamental [5,9], reduced-form [10,11], statistical [12–14], and computational intelligence (CI) models [15–17]. Compared with the other four traditional models, CI models are regarded as state-of-the-art techniques, and their superior performance contributes to their prevalence in EPF in recent years. In particular, deep neural networks (DNNs) have gradually become the most avant-garde CI approach in other disciplines [18–20] and entered the scientific research field related to EPF.

DNNs are often categorised into three main classes: feed-forward neural networks (FNNs), recurrent neural networks (RNNs), and convolutional neural networks (CNNs). Different types of DNNs are used to solve different problems. For time series

* Corresponding author.
E-mail addresses: wei.n.li@ntnu.no (W. Li), denis.becker@ntnu.no (D.M. Becker).

https://doi.org/10.1016/j.energy.2021.121543
0360-5442/

Nomenclature

β	Coefficients of independent variables
$\hat{y}$	Predicted output
λ	Tuning parameter
μ	Expected electricity exchange capacity
ω	Constant that determines the significance of $\overrightarrow{v}$
$\otimes$	Element-wise product
$\overrightarrow{g}$	Global current best solution
$\overrightarrow{p}$	Particle's past best solution
$\overrightarrow{v}$	Velocity of the particle
$\mathbf{\Phi}$	Kernel function
φ	Activation function
ρ	PC coefficient
σ_X	Standard deviation of X
σ_Y	Standard deviation of Y
σ_{FD}	Cross-border flow deviation
$\mathbf{w}$	Parameter vector
Tanh	Hyperbolic tangent function
ϵ	Error variable
ξ, ξ_i^*	Slack variable
ζ	Sigmoid function
b	Bias vector
C	Regularisation constant
c	Cell state vector
c_1	Constant that determines the significance of $\overrightarrow{p}$
c_2	Constant that determines the significance of $\overrightarrow{g}$
cov	Covariance
$d^{F1,F2}$	Loss differential
e	Forecast error
f	Forget gate's activation vector
$g(*)$	Loss function
h	Hidden state vector
i	Input/update gate's activation vector
o	Output gate's activation vector
r_1, r_2	Random number from the interval [0, 1]
W	Weight matrix
w^1	Weight vector between the input and the hidden layer
w^2	Weight vector between the hidden layer and the output
X, Y	Variable
x	Input variable
y	Output variable

Abbreviations

CNNs	Convolutional neural networks
ConvLSTM	Convolutional LSTM
DM	Diebold-Mariano
DNNs	Deep neural networks
ELM	Extreme learning machine
EPF	Electricity price forecasting
FNNs	Forward neural networks
GA	Genetic algorithm
GRUs	Gated recurrent units
LSTM	Long short-term memory
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MSE	Mean squared error
NARMAX	Moving Average with eXogenous Input
PC	Pearson's Correlation
PSO	Particle swarm optimisation
RMSE	Root mean squared error
RNNs	Recurrent neural networks
SHAP	SHapley Additive exPlanations
SMAPE	Symmetric mean absolute percentage error

prediction, RNNs have achieved superior performance by building extra mappings to hold relevant information from past inputs. The long short-term memory (LSTM) and gated recurrent units (GRUs) are important variants of this kind of network, which overcome the vanishing gradient problem of RNNs [21]. Compared with GRUs, LSTM is more accurate on the dataset using long sequences. Due to the superiority of LSTM in time series forecasting, researchers have gradually paid attention to its application in EPF [22–24]. However, as with other DNNs, when LSTM models are applied to high-dimensional data, a critical issue occurs, known as the curse of dimensionality [25]. This means that, with a large amount of features,[1] the performance of LSTM will degrade because of overfitting [26]. Thus, LSTM cannot be employed directly for electricity price prediction with a large number of features as input under consideration of market coupling. While some researchers have attempted to involve explanatory variables from integrated markets to make a prediction of electricity price by neural networks [27–29], no existing research has investigated the state-of-the-art LSTM-based deep neural networks for this purpose. Besides, some research starts to pay attention to the influence of the market integration on Nord Pool [30–32]. However, efficient ways to utilise the ever-growing information from the electricity market integration for the Nordic EPF have yet to be explored.

Typically, feature selection is an efficient way to avoid the curse of dimensionality. It is the process of selecting a subset of relevant attributes in the dataset when developing a predictive model. It can reduce the computation time, improve model prediction performance, and help to get a better understanding of the dataset [33]. The ideal feature selection is to search the space of all variable subsets with an algorithm, which is impractical except for quite small sized feature spaces. However, as the space of variables subset grows exponentially with the number of variables, heuristic search methods are commonly used to search for an optimal subset [34]. The current research on feature selection algorithms can be categorised as filter, wrapper, and embedded methods [33]. In particular, the filter methods use a proxy measure to estimate a feature subset before training a prediction model. Pearson's Correlation (PC) is a typical indirect assessing measure for the regression problem [35]. In contrast, the wrapper methods evaluate selected feature subsets by employing a predictive model directly. Each subset is used to train a new forecasting model, and the optimisation method is used to search for the best performing model in the process of feature selection. The embedded methods can implement an automatic feature selection in the process of estimating the parameters of predictive models. This means this catch-all group of techniques performs the process of feature

[1] In machine learning, features are individual independent variables as input in a model.

selection during the training of the model. The particle swarm optimisation combined with the extreme learning machine method (PSO-ELM) and genetic algorithm combined with the extreme learning machine method (GA-ELM) are two typical wrapper-based methods. They have been widely used for various feature selection problems [36–41]. Guyon et al. [42] proposed another popular wrapper approach, known as recursive feature elimination combined with support vector machine for regression (RFE-SVR). The Lasso regression method is one of the most popular embedded feature selection methods proposed by Tibshirani [43].

1.1. Contributions

To the best of our knowledge, no existing study considers how to apply LSTM models in an integrated market EPF and detect the impact of the features from cross-border markets on EPF. To fill this scientific gap, we propose three hybrid architectures of LSTM-based deep learning predictive models combined with advanced feature selection algorithms: the two-step hybrid architecture, the autoencoder hybrid architecture, and the two-stage hybrid architecture. Different feature selection methods have different selection mechanisms, which will lead to different sets of selected features. To explore the influence of different feature selections on LSTM-based EPF, we employed five feature selection algorithms, PC, PSO-ELM, GA-ELM, RFE-SVR, and the Lasso regression method, in the case study of Nord Pool and its neighbouring, interconnected countries. The main contributions of this study are as follows:

1. We compare and analyse the forecasting performance of the proposed models in the case study of the Nord Pool system price forecasting, considering six integrated markets (sixty-two features). The results indicate that the cross-border markets influence the Nordic electricity price formation. As the rapid market coupling development in EU, we show that it is necessary to consider cross-border information for EPF in future studies.
2. We introduce three architectures of hybrid LSTM-based deep neural networks for EPF and conclude that different feature selection algorithms yield divergent subsets of features, which, in turn, affect the prediction accuracy of the proposed LSTM models. In addition, the results show that hybrid models are an efficient way to deal with the ever-growing information and obtain accurate prediction results in cross-border markets.
3. We employ a game theoretical approach (SHapley Additive exPlanations) to explore the relevance of various cross-border features in EPF. The analysis of Shapley values increases the transparency of the prediction and provides advice for policy makers and market participants.

The remainder of this paper is organised as follows. Section 2 describes the dataset used in this research. In Section 3, we present the methodology. Section 4 describes the model training and introduces evaluation criteria applied in the empirical study. Section 5 reports the forecasting results of the implemented models. Finally, Section 6 concludes the paper and proposes future research developments.

2. Data description

The Nordic system price is the central reference price in the Nordic electricity market. It is used as a settlement price for the derivatives market. Each hourly system price is calculated by Nord Pool based on all bids and offers posted in Nordic bidding zones, which is referred to as a market-clearing price, without taking into account any congestion restrictions. The daily system price represents the arithmetic average of the 24 hourly prices. This paper discusses and evaluates several hybrid LSTM-based approaches for the prediction of the Nordic hourly and daily system prices.

Previous empirical research on the prediction of electricity prices has considered information from both price and supply/demand sides. To find out what matters when predicting the day-ahead Nordic system price in coupling markets, we also included the electricity exchange between Nord Pool and its integrated countries. The Russian electricity market is excluded because it differs significantly from European models. To consider the influence of the correspondence between electricity flow and capacity, we introduced a new daily feature, namely the cross-border flow deviation. It can be calculated as $\sigma_{FD} = \sqrt{\sum_{i=1}^{N}(X_i - \mu_i)^2}/N$, where X_i is the hourly electricity flow, μ_i is the hourly expected exchange capacity, and N stands for 24 h.

In summary, we consider eight categories of input features: day-ahead price, production, production prognosis, consumption, consumption prognosis, currency exchange rate, cross-border electricity flow, and flow deviation. The first five are the basic features from local markets for predicting electricity price. Some historical/predictive information, such as weather and human social activities, does not directly impact electricity price but influences the supply/demand for electricity, incorporated in those five fundamental variables. The last three are the features spawned from cross-border trade.

2.1. Data

We collected data from the Nord Pool,[2] Thomson Reuters Eikon,[3] and Entsoe.[4] The available time series ranges from 01/01/2015 to 31/12/2019. Nord Pool provides cross-border transmissions with Germany (DE), the Netherlands (NL), Lithuania (LT), Estonia (EE), Poland (PL), and Russia (RU). The map in Fig. 1 shows both the Nord Pool markets as well as transmissions (black dashed lines) between the Nord Pool and its coupling bidding areas. There are five bidding zones in Norway (NO1, NO2, NO3, NO4, and NO5), four in Sweden (SE1, SE2, SE3, and SE4), and two in Denmark (DK1 and DK2), and one in Finland (FI). Since the transmissions between DK1 and NL started at 01/09/2019, the data series is not sufficient for the application of deep learning models. Besides, the electricity exchange between SE4 and LT started at 09/12/2015. Therefore, the entire available dataset employed in this study ranges from 09/12/2015 to 31/12/2019. The features included in the dataset are shown in Table 1. The hourly data is converted into the daily data by the arithmetic average (e.g., price) or the aggregate (e.g., flow).

2.2. Cross-border electricity transmission

Fig. 2 shows the electricity exports from Germany, the Netherlands, Lithuania, Poland, and Russia in 2019.[5] The exports to the Nord Pool comprised 16.03 % of the whole exports from these coupling countries. In Fig. 3, we can see that the electricity exports of the Nord Pool comprised 4.82 % of its total production in 2019. The EU aims to achieve 15 % interconnection capacity in 2030 for each EU country [44].

[2] Nord Pool: https://www.nordpoolgroup.com/.
[3] Thomson Reuters Eikon: https://eikon.thomsonreuters.com/.
[4] Entsoe: https://transparency.entsoe.eu/.
[5] Fraunhofer ISE provides the electricity exchange data of Germany/Europe: https://www.energy-charts.de/.

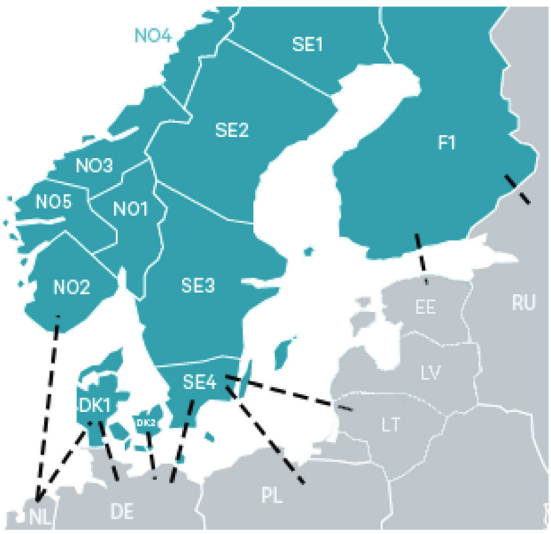


Fig. 1. Overview of the Nord Pool market coupling.

3. Methodology

3.1. LSTM

The LSTM architecture was initially introduced by Hochreiter and Schmidhuber [45] and has since been enhanced by other researchers to achieve better performance [46–48]. An LSTM network is a special kind of recurrent neural network that is capable of learning long-term dependencies. Unlike simple RNNs, an LSTM network has built-in mechanisms that control how information is memorised or abandoned throughout time. The architecture of the LSTM network is shown in Fig. 4 and is defined by the following system of equations [49]:

$$f_t = \zeta\left(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f\right) \quad (1)$$

$$i_t = \zeta(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (2)$$

$$o_t = \zeta(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_{t-1} + b_o) \quad (3)$$

$$c_t = f_t \otimes c_{t-1} + i_t \otimes \mathrm{Tanh}(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (4)$$

$$h_t = o_t \otimes (c_t) \quad (5)$$

where f_t, i_t, o_t, c_t, and h_t indicate the values of the forget gate state, input gate state, output gate state, memory cell, and hidden state at time t in the sequence, respectively. ζ and Tanh are the sigmoid function and hyperbolic tangent function, W and b are the weight matrix and bias vector, and $\otimes$ denotes the element-wise product.

3.2. Architectures of hybrid models

Typically, there are three hybrid architectures for EPF when working with LSTM. The first architecture consists of two steps, as shown in Fig. 5. The first step includes data processing and feature selection, and the second step contains training of the predictive models and making predictions.The second architecture can be referred to as an autoencoder model. Here, the input data will be turned into a compressed representation rather than specifically showing which features are selected, as shown in Fig. 6. The third combines the two aforementioned architectures, and it is referred to as two-stage feature selection. In this architecture, the explanatory variables will be selected by some feature selection method in

Table 1
The features included in the dataset.

Feature	Description (Units)	Data Source
F1	System Day-ahead price 1-Lag (EUR/MWh)	Nord Pool
F2	SE1 Day-ahead price (EUR/MWh)	Nord Pool
F3	SE2 Day-ahead price (EUR/MWh)	Nord Pool
F4	SE3 Day-ahead price (EUR/MWh)	Nord Pool
F5	SE4 Day-ahead price (EUR/MWh)	Nord Pool
F6	FI Day-ahead price (EUR/MWh)	Nord Pool
F7	DK1 Day-ahead price (EUR/MWh)	Nord Pool
F8	DK2 Day-ahead price (EUR/MWh)	Nord Pool
F9	NO1 Day-ahead price (EUR/MWh)	Nord Pool
F10	NO2 Day-ahead price (EUR/MWh)	Nord Pool
F11	NO3 Day-ahead price (EUR/MWh)	Nord Pool
F12	NO4 Day-ahead price (EUR/MWh)	Nord Pool
F13	NO5 Day-ahead price (EUR/MWh)	Nord Pool
F14	EE Day-ahead price (EUR/MWh)	Nord Pool
F15	LT Day-ahead price (EUR/MWh)	Nord Pool
F16	PL Day-ahead price (PLN/MWh)	Thomson Reuters Eikon
F17	DE Day-ahead price (EUR/MWh)	Thomson Reuters Eikon
F18	NL Day-ahead price (EUR/MWh)	Thomson Reuters Eikon
F19	Nordic production (MWh)	Nord Pool
F20	EE production (MWh)	Nord Pool
F21	LT production (MWh)	Nord Pool
F22	PL production (MWh)	Entsoe
F23	DE production (MWh)	Entsoe
F24	NL production (MWh)	Entsoe
F25	Nordic production prognosis (MWh)	Nord Pool
F26	EE production prognosis (MWh)	Nord Pool
F27	LT production prognosis (MWh)	Nord Pool
F28	PL production prognosis (MWh)	Entsoe
F29	DE production prognosis (MWh)	Entsoe
F30	NL production prognosis (MWh)	Entsoe
F31	Nordic consumption (MWh)	Nord Pool
F32	EE consumption (MWh)	Nord Pool
F33	LT consumption (MWh)	Nord Pool
F34	PL consumption (MWh)	Entsoe
F35	DE consumption (MWh)	Entsoe
F36	NL consumption (MWh)	Entsoe
F37	Nordic consumption prognosis (MWh)	Nord Pool
F38	EE consumption prognosis (MWh)	Nord Pool
F39	LT consumption prognosis (MWh)	Nord Pool
F40	PL consumption prognosis (MWh)	Entsoe
F41	DE consumption prognosis (MWh)	Entsoe
F42	NL consumption prognosis (MWh)	Entsoe
F43	EUR/NOK	Nord Pool
F44	EUR/SEK	Nord Pool
F45	EUR/DKK	Nord Pool
F46	EUR/PLN	Thomson Reuters Eikon
F47	NO2 ↔ NL flow (MWh)	Nord Pool
F48	DK1 ↔ DE flow (MWh)	Nord Pool
F49	DK2 ↔ DE flow (MWh)	Nord Pool
F50	SE4 ↔ DE flow (MWh)	Nord Pool
F51	SE4 ↔ PL flow (MWh)	Nord Pool
F52	SE4 ↔ LT flow (MWh)	Nord Pool
F53	FI ↔ EE flow (MWh)	Nord Pool
F54	FI ↔ Russia flow (MWh)	Nord Pool
F55	NO2 ↔ NL flow deviation	Calculation
F56	DK1 ↔ DE flow deviation	Calculation
F57	DK2 ↔ DE flow deviation	Calculation
F58	SE4 ↔ DE flow deviation	Calculation
F59	SE4 ↔ PL flow deviation	Calculation
F60	SE4 ↔ LT flow deviation	Calculation
F61	FI ↔ EE flow deviation	Calculation
F62	FI ↔ Russia flow deviation	Calculation

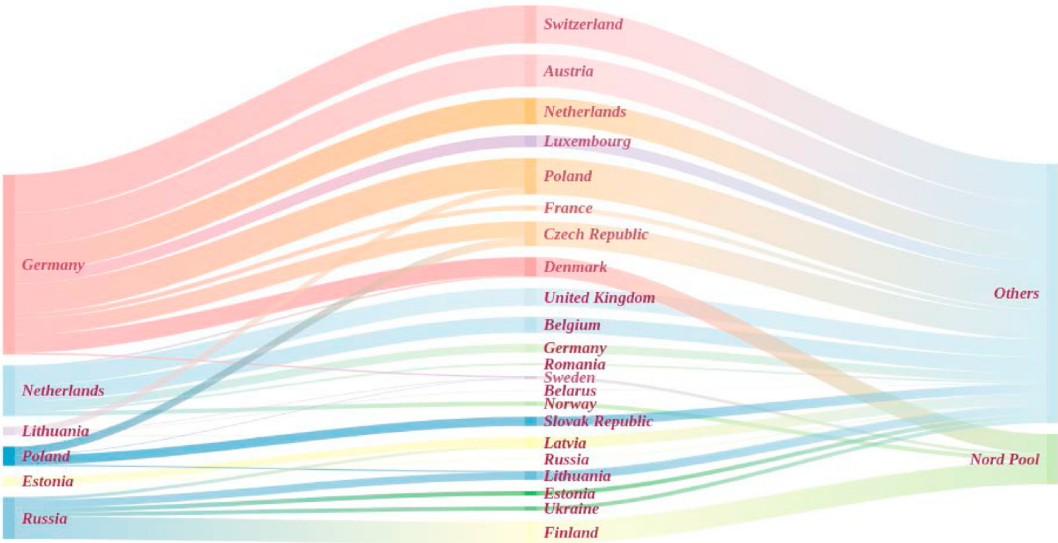


Fig. 2. The electricity cross-border transmission from the coupling countries to Nord Pool.

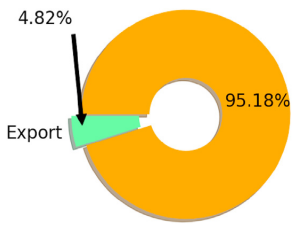


Fig. 3. The percentage of the Nord Pool production for exporting.

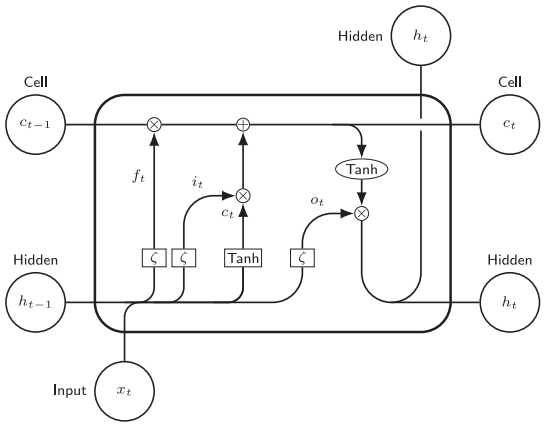


Fig. 4. LSTM cell.

the first stage. The selected features will then become the input for the autoencoder models in the second stage. Fig. 7 shows this architecture.

3.3. Feature selection methods

3.3.1. PC

The PC coefficient is a statistic used to measure the linear relationship between two data samples. Given two variables (X, Y), the formula of the PC coefficient ρ is given by the following:

$$\rho(X,Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \tag{6}$$

where cov is the covariance, σ_X is the standard deviation of X, and σ_Y is the standard deviation of Y.

3.3.2. PSO-ELM and GA-ELM

PSO-ELM and GA-ELM are wrapper-based hybrid methods. ELM is a single hidden layer feedforward neural network. Its fast training [50] contributes to the popularity of its employment as a predictive model in wrapper-based feature selection [51–53]. The output of ELM is calculated as follows:

$$F_L(x) = \sum_{i=1}^{L} w_i^2 \varphi\left(w_i^1 x_j + b_i\right), j = 1, \ldots, N \tag{7}$$

where L is the number of hidden units, N is the number of training samples, w^2 is the weight vector between the hidden layer and the output, w^1 is the weight vector between the input and the hidden layer, $\varphi(*)$ denotes an activation function, b is a bias vector, and x is the input vector.

PSO and GA are different types of optimisation algorithms, which provide the subsets of features as the input to the ELM to detect the optimal feature selection. The basic idea of PSO is that a swarm of particles moves through the search space. The movement of each particle is guided by its own known best-position and the entire swarm's known best position. PSO performs the search for the optimum by iteratively updating the velocities of the particles in the swarm [54]. The GA is a search metaheuristic that was inspired by Darwin's theory of natural selection. In general, GAs will search for

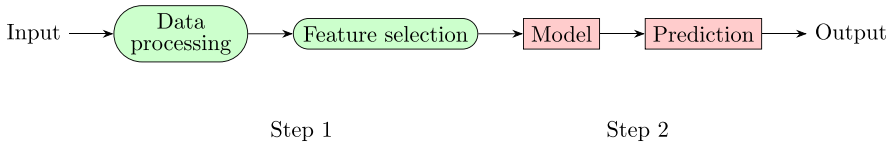


Fig. 5. The flowchart of a two-step hybrid model. The green nodes stand for the first step, and the red nodes stand for the second step. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

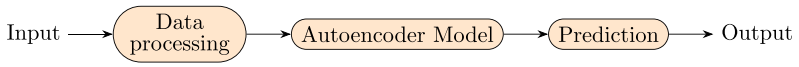


Fig. 6. The flowchart of an autoencoder hybrid model. The orange nodes stand for the autoencoder process. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

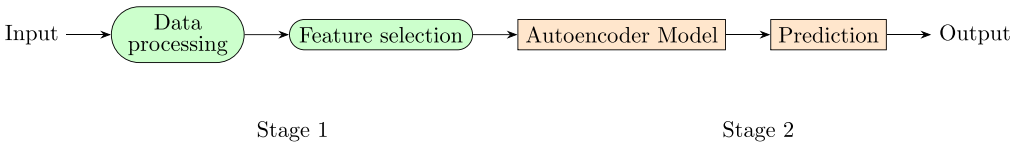


Fig. 7. The flowchart of a two-stage hybrid model. The green nodes stand for the first stage, and the orange nodes stand for the second stage. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

the optimal solution from a set of possible solutions, called a population. A solution is referred to as a chromosome or an individual. These chromosomes evolve over a number of generations by recombination (cross-over) and mutation [55]. The detailed introduction of the methods can be found in Appendix A.1 and A.2.

3.3.3. RFE-SVR

RFE-SVR is another wrapper-based feature selection method. The core idea of this algorithm is to search for the best subset of features by starting with all features and discarding the less important features. In particular, the RFE algorithm operates with SVR to perform feature selection and regression simultaneously. SVR performs well in high dimensionality space [56]. The detailed explanation of SVR is in Appendix A.3.

3.3.4. Lasso regression

The Lasso regression aims to increase the prediction accuracy of regression models by adding a penalty $\lambda\sum_{j=1}^{n}|\beta_j|$ to the loss function. This means that instead of minimising a loss function, $\sum_{i=1}^{m}(y_i-\sum_{j=1}^{n}x_{ij}\beta_j)^2$, the loss function becomes $\sum_{i=1}^{m}(y_i-\sum_{j=1}^{n}x_{ij}\beta_j)^2+\lambda\sum_{j=1}^{n}|\beta_j|$, where y is the vector of the dependent variable, x denotes independent variables, the β are the corresponding coefficients. The algorithm has the advantage that it shrinks some of the less critical coefficients of features to zero. Therefore, it removes less relevant features.

3.4. Autoencoder model

An autoencoder is typically a neural network that aims to filter and compress the representation of its input, which consists of two components: an encoder and a decoder, shown in Fig. 8. The encoder typically accepts a set of input data and compresses the information into an intermediate vector. The decoder is typically a predictive model. In our case, a decoder is an LSTM network, and the encoders are LSTM, CNN, and convolutional layers, as described in the following.

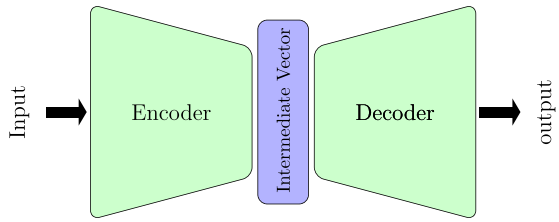


Fig. 8. Structure of the Encoder-Decoder model.

3.4.1. LSTM-LSTM Encoder-Decoder model

In an LSTM-LSTM Encoder-Decoder model, an LSTM model is used as the encoder to process the raw input time series and to transform it into an intermediate vector. LSTM is capable of extracting the complex dynamic information within the temporal input series and filtering useful information from long input series via internal memory.

3.4.2. CNN-LSTM Encoder-Decoder model

In a CNN-LSTM Encoder-Decoder model, a CNN is the encoder to filter the input data. CNNs were originally and successfully used to process the image input data in image recognition tasks [57] or the sequence of input data in natural language processing problems [58]. The convolutional layers are usually followed by a pooling layer, which extracts information from the convolved features and produces a lower dimensional output. Then, the output values are flattened into a long intermediate vector representation.

3.4.3. Convolutional LSTM Encoder-Decoder model

The computational mechanism of the convolutional LSTM (ConvLSTM) is similar to that of CNN-LSTM [59]. Unlike the CNN-LSTM, where the CNN model generates the input for the LSTM model, in the ConvLSTM model, the LSTM neural network processes the extracted information directly from preceding convolutional layers.

4. Experimental details

In this section, we introduce the concepts and methods employed in the process of training and evaluation of the constructed models. This section includes five parts. First, statistical performance measures are specified for evaluation and comparison. Second, theory and preliminary work for training models are explained. Third, configuration parameters are set for feature selection algorithms and LSTM-based models. Fourth and fifth, the benchmark model and feature explanation method used in the empirical study are introduced. They are necessary preparations for conducting the experiments.

4.1. Statistical performance measures

4.1.1. Evaluation metrics

In this paper, we employ several indicators to evaluate the accuracy of predictions: the mean absolute error (MAE), the root mean squared error (RMSE), the mean absolute percentage error (MAPE), and the symmetric mean absolute percentage error (SMAPE) as the model estimator. They are commonly adopted in EPF research [1]. Given a predicted output vector, $\hat{y_k} = [\hat{y_1}, .., \hat{y_N}]$, and a real output vector, $y_k = [y_1, .., y_N]$, the MAE, RMSE, MAPE, and SMAPE can be calculated as follows:

$$\mathrm{MAE} = \frac{1}{N}\sum_{k=1}^{N}|y_k - \hat{y_k}| \tag{8}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(y_k - \hat{y_k})^2} \tag{9}$$

$$\mathrm{MAPE} = \frac{100}{N}\sum_{k=1}^{N}\frac{y_k - \hat{y_k}}{y_k} \tag{10}$$

$$\mathrm{SMAPE} = \frac{100}{N}\sum_{k=1}^{N}\frac{|y_k - \hat{y_k}|}{(|y_k| + |\hat{y_k}|)/2} \tag{11}$$

4.1.2. Diebold-Mariano test

The metrics for assessing the forecasting accuracy mentioned above cannot guarantee that the observed difference from two predictive models is statistically significant. In this context, the Diebold-Mariano (DM) test is typically used for evaluating the performance of two models [60,61]. Given the actual values of a time series $[y_t; t = 1, \ldots, T]$, two forecasts from two models, $[\hat{y}_{1t}; t = 1, \ldots, T]$ and $[\hat{y}_{2t}; t = 1, \ldots, T]$, and the associated forecast errors, $e_{1t} = \hat{y}_{1t} - y_t$ and $e_{2t} = \hat{y}_{2t} - y_t$, the DM test defines the loss differential between the two forecasts by the following:

$$d_t^{F1,F2} = g(e_{1t}) - g(e_{2t}) \tag{12}$$

where $g(*)$ stands for loss function. In a one-sided DM test, the hypotheses is the following:

$$\begin{aligned} H_0 &: \mathbb{E}\left(d_t^{F1,F2}\right) \geq 0, \\ H_1 &: \mathbb{E}\left(d_t^{F1,F2}\right) < 0. \end{aligned} \tag{13}$$

A one-sided DM test is used to detect whether F2 is better than F1. If H_0 is rejected, the test suggests that the accuracy of F1 is, statistically, significantly better than F2.

The complementary one-sided DM test can be expressed as follows:

$$\begin{aligned} H_0 &: \mathbb{E}\left(d_t^{F1,F2}\right) \leq 0, \\ H_1 &: \mathbb{E}\left(d_t^{F1,F2}\right) > 0. \end{aligned} \tag{14}$$

If H_0 is rejected, the test suggests that the accuracy of F2 is, statistically, significantly better than F1. In this study, we employ a one-sided DM test to assess the forecasting performance of the proposed models. We chose $d_t^{F1,F2} = |e_{1t}| - |e_{2t}|$ as the loss differential.

4.2. Model training

4.2.1. Walk forward nested cross-validation

To avoid over-fitting, it is common to include a validation set to evaluate the generalisation ability of the training model. The cross-validation is referred to as a method for tuning the hyper-parameters and producing robust measurements of model performance. In the study of Varma and Simon [62], a nested cross-validation procedure was introduced, which considerably reduced the bias and provided an almost unbiased estimate of the true error. Because new observations become available over time, in time series modelling, we implemented a walk forward nested cross-validation in which the forecast rolls forward in time. More specifically, we successively considered each day as the test set and assigned all previous data to the training set (Outer loop). The training set is split into a training subset and a validation set. The validation set data comes chronologically after the training subset (Inner loop). Walk forward validation involves moving along the time series one time step at a time. The process requires multiple models to be trained and evaluated, but the additional computational cost will provide a more robust estimate of the expected performance of the predictive model on unseen data. It is shown in Fig. 9.

4.2.2. Data division

We divided the whole database into two subsets: a training set and a test set. The training set includes a training subset and a validation subset, as shown in the dashed box in Fig. 9. We initially apportioned the data set into training, validation, and test sets, with an 80-10-10 split. The magnitude of the test and validation set is anchored during the walk-forward test.

4.2.3. Data processing

For neural network model training, the input data is usually normalised to the intervals [0,1]. This is not only done because the

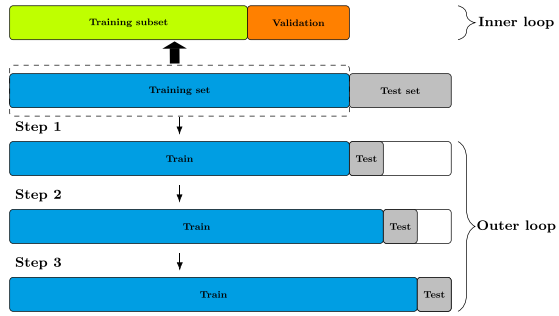


Fig. 9. Walk forward nested cross-validation.

normalised data will require less time to train, but the prediction performance will also increase. In addition, we linearly interpolate the missing data and eliminate duplicates due to daylight saving.

4.2.4. Ten experiments

Training algorithms for deep learning models have usually required the initialisation of the weights of neural networks from which to begin the iterative training [63]. The random initial conditions for an LSTM network can result in different performances each time a given configuration is trained. Thus, we employed ten experiments for each model to reduce the impact of the variability on performance evaluation. Models were evaluated after taking the average of the experiments.

4.3. Model configuration parameters

4.3.1. Parameters of feature selection

The feature selection stopping criterion varies by algorithm, which is controlled by the parameters of models. The applied configuration of PSO was [c_1: 0.5, c_2: 0.3, ω: 0.7], and the stop condition is satisfied after 10,000 iterations. For GA, the crossover possibility and mutation possibility were set to 0.5 and 0.2, respectively. The population size was 100, and the maximum number of generations was 10,000. On the basis of the predictive ELM, the amount of the selected features by PSO-ELM and GA-ELM was automatically set to 30. For the sake of input consistency, the magnitude of the selected features of the rest of the models was set to 30 as well. For the PC method, we ranked all features attributable to the correlation coefficients and selected the first 30 features. In terms of RFE-SVR, we ranked features by importance, discarded the least important features, and refit the model until 30 features remained. The regularisation parameter, λ, in Lasso regression was 0.02.

4.3.2. Network hyperparameters

Our study aimed to investigate the applications and impacts of different types of feature selection methods in a predictive LSTM architecture. We used a coherent configuration of a specific LSTM model for comparison and did not perform an extensive hyperparameter optimisation to search for the optimal configuration. After an inexhaustive grid search, we constructed our prediction model from an LSTM model with a single hidden layer of 300 units, followed by a fully connected dense layer with 100 neurons that preceded the output layer. The LSTM encoder has a hidden layer with 300 units. In the CNN-LSTM Encoder-Decoder model, the CNN encoder has two convolutional layers, with 96 units to amplify any salient features, followed by a max-pooling layer. In the ConvLSTM Encoder-Decoder model, the encoder is a convolutional layer with 64 units. The input sequence length is 14 days (2 weeks, commonly used in EPF). The optimiser is the Adam algorithm, and the loss function is Mean Squared Error (MSE).

4.4. Benchmark model

Among the traditional methods, the statistical models perform best for EPF. Thus, we select the Nonlinear AutoRegressive Moving Average with eXogenous Input (NARMAX) model as the benchmark (trained with the optimal structure) for our case study. This statistical model is widely used in energy price forecasting to handle multiple nonlinear inputs [32,64]. The equation is represented as the following:

Table 2
The proposed models.

Mode	Category	Model Explanation
M0	Benchmark	NARMAX model
M1	Filter method	PC-LSTM model
M2	Wrapper method	PSO-ELM-LSTM model
M3	Wrapper method	GA-ELM-LSTM model
M4	Wrapper method	RFE-SVR-LSTM model
M5	Embedded method	LASSO-LSTM model
M6	Autoencoder method	LSTM-LSTM Encoder-Decoder model
M7	Autoencoder method	CNN-LSTM Encoder-Decoder model
M8	Autoencoder method	CovLSTM Encoder-Decoder model
M9	Two-stage method	PC-LSTM-LSTM Encoder-Decoder model
M10	Two-stage method	PSO-ELM-LSTM-LSTM Encoder-Decoder model
M11	Two-stage method	GA-ELM-LSTM-LSTM Encoder-Decoder model
M12	Two-stage method	RFE-SVR-LSTM-LSTM Encoder-Decoder model
M13	Two-stage method	LASSO-LSTM-LSTM Encoder-Decoder model

$$y(t) = F^{\ell}\big[y(t-1),\ldots,y(t-N_y),x(t),\ldots,x(t-N_x),\ \epsilon(t-1),\ldots\epsilon(t-N_\epsilon)\big] + \epsilon(t) \tag{15}$$

where $x(t)$ is the input and $y(t)$ is the output time-series; $\epsilon(t)$ is the uncertainties and possible noise; N_u, N_y, and N_ϵ are the input, output, and prediction error lags, respectively; and F^{ℓ} is a nonlinear function.

4.5. Feature explanation method

In this study, we used SHapley Additive exPlanations (SHAP) values to interpret the impact of certain values of a given feature on the expected price prediction. SHAP[6] is a theoretic game method to explain the output of machine learning models [65–67]. The Shapley value is used to assess the feature relevance relative to the expectation of the output [68]. In particular, a Kernel SHAP is used for explaining an optimal SVR model obtained by grid-search on the dataset.

5. Results

In this section, we report the empirical results obtained by the application of the introduced models. For similarity of presentation, the list of models and their acronyms are shown in Table 2.

5.1. Analysis of empirical results

The results of the feature selection are shown in Table 3. Overall, it can be observed that different selection mechanisms lead to different selections. From Table 3, we can see that M1 selects all the day-ahead prices. It is not surprising that the day-ahead prices from different bidding areas are more relevant to the Nord Pool system price than other feature variables. However, the over-selection results in information redundancies. Some researchers have recognised that the redundancy among features decreases the model's performance [69–71]. Compared to M1, the wrapper-based methods, M2 and M3, eliminate several price variables rather than the other categories of variables. It is worth noticing the two methods does not select the lag system price (F1), which is commonly used in time series, given their short-term autoregressive nature. As introduced before, PSO-ELM and GA-ELM are widely used in research. However, the optimisation methods, such as PSO and GA, have the problem of trapping in local optima. Although re-setting and experimenting can increase the chance of

[6] The Python package SHAP is available at https://github.com/slundberg/shap.

Table 3
The results of feature selection.

Feature	Feature selection model				
	M1	M2	M3	M4	M5
F1	✓	×	×	✓	✓
F2	✓	✓	✓	✓	✓
F3	✓	×	×	✓	✓
F4	✓	✓	×	✓	×
F5	✓	×	×	×	×
F6	✓	×	×	×	×
F7	✓	×	×	×	×
F8	✓	✓	×	×	×
F9	✓	×	×	✓	×
F10	✓	×	×	✓	✓
F11	✓	✓	×	✓	✓
F12	✓	×	✓	✓	✓
F13	✓	✓	✓	✓	✓
F14	✓	×	✓	×	×
F15	✓	×	×	×	×
F16	✓	×	×	×	×
F17	✓	×	✓	×	×
F18	✓	✓	×	✓	✓
F19	×	✓	×	✓	×
F20	×	✓	✓	✓	×
F21	×	×	✓	×	×
F22	×	✓	✓	✓	×
F23	×	✓	×	✓	×
F24	✓	✓	✓	×	×
F25	×	×	×	✓	✓
F26	×	×	✓	✓	✓
F27	✓	✓	✓	×	✓
F28	×	✓	✓	×	×
F29	×	×	×	✓	✓
F30	×	✓	×	×	✓
F31	×	✓	×	✓	×
F32	×	✓	✓	✓	✓
F33	✓	✓	×	×	×
F34	✓	×	✓	✓	✓
F35	×	×	×	✓	✓
F36	×	✓	✓	✓	✓
F37	×	×	×	✓	×
F38	✓	✓	✓	✓	×
F39	✓	✓	✓	✓	✓
F40	×	×	✓	×	×
F41	✓	✓	×	✓	✓
F42	×	×	×	×	×
F43	✓	×	✓	✓	×
F44	✓	✓	✓	✓	✓
F45	✓	×	✓	×	×
F46	×	✓	×	×	✓
F47	×	×	×	×	×
F48	×	×	×	✓	✓
F49	×	×	✓	×	✓
F50	×	✓	✓	×	✓
F51	×	✓	×	×	✓
F52	×	✓	✓	×	×
F53	×	×	×	✓	✓
F54	×	×	✓	×	✓
F55	×	×	×	×	×
F56	×	×	✓	×	×
F57	×	✓	✓	×	×
F58	×	×	✓	×	✓
F59	✓	✓	✓	×	✓
F60	✓	✓	✓	×	×
F61	×	✓	×	×	✓
F62	×	×	×	×	×

Note: ✓ denotes that the feature is selected.
× denotes that the feature is not selected.

avoiding traps, when dealing with high dimensional data sets, the optimisation methods still cannot guarantee they will find a global optimum solution, and they are not suitable for all cases [72]. The straightforward concept and fast computation of ELM contributed to its widespread application in an exhaustive grid search, but it does not consider the sequential relationships in time series data, as with other traditional neural networks. These could be the reasons why the two methods eliminate the lag system price as an input. The other wrapper model M4, selects various types of features, and eliminates less price features compared to M2 and M3. It is interesting to note that M4 does not pick up any features from the cross-border flow deviation. For the Lasso regression method, M5, we can summarise that it diversely chooses features such as M4 but puts more emphasis on electricity transmission.

To evaluate the statistical significance in the difference of predictive accuracy, one-side DM tests, as defined in section 4.1.2, were applied, and the results are shown in Table 4. Table 5 exhibits the performance comparison of all the models in terms of SMAPE. As expected, the proposed LSTM models are overwhelmingly better than the benchmark statistical model, M0. The superior performance of deep learning models to statistical models has been recognised by numerous studies [22,73]. Moreover, the performances of LSTM models in ten experiments are depicted in Fig. 10, measured in terms of SMAPE. As seen in Fig. 10, of all the models, M4, M5, and M6 perform better than the others. The statistical details of the model performance in ten experiments are listed in Appendix Tables A.7, A.8, A.9, and A.10 by means of MAD, RMSE, MAPE, and SMAPE. Based on the analysis of feature selection, we conclude that M4 and M5, i.e., the minimum redundancy maximum relevance algorithms, perform better than the others. The results are consistent with the observations from other literature [74,75] that an elimination of the redundant and less relevant features increases the performance of models. In addition, some researchers have attempted to introduce the CNN-LSTM model and show its excellent performance in the energy field [76,77]. However, we found that M6 performs better than M7 and M8. This means LSTM-LSTM is a better autoencoder structure than CNN-LSTM and ConvLSTM for EPF. The results are reasonable because convolutional neural networks (CNNs or ConvNets) were originally designed for image recognition and classification, while recurrent neural networks (LSTM) are for sequence and time series prediction.

We show the comparison of the SMAPE of the two-step and the two-stage models in Fig. 11. From this figure, it can be seen that M4 has been improved by applying LSTM-LSTM as predictors. We used a one-sided DM test to detect whether the two-stage LSTM models were statistically better than two-step LSTM-LSTM models. The results are shown in Table 6. The superior features selection from M4 provides the possibility for autoencoder models to further process the selected features to obtain more meaningful information.

Additionally, we detected the forecasting performance for 24 hourly system prices. Figs. 12–14 show the results for the three peak hours: H8 (07–08), H12 (11–12) and H18 (17–18), respectively, measured in terms of SMAPE. We observed that the feature selections influence the forecasting accuracy and the models M4 and M5 are still relatively stable, performing better than other models. Indeed, the RFE-SVR and Lasso regression feature selection methods are applied to various areas in energy finance and achieve good performance for improving forecasting accuracy [78–81].

5.2. Analysis of feature impact

Fig. 15 shows the ranking of features and their impacts on the predicted price in terms of the selected features of the RFE-SVR model (M4), which is the model with the best performance. From Fig. 15a, we can observe that the features from supply/demand sides are more important than the other features. In particular, production and consumption and their prognosis in the Nordic and German markets are prioritised by the model. The significant impact from the German market can be explained by the fact that

Table 4
The results of the one-sided DM test.

F1	F2								
	M0	M1	M2	M3	M4	M5	M6	M7	M8
M0		5.83***	4.99***	3.90***	6.65***	7.09***	6.60***	4.61***	4.27***
M1	−5.83***		−1.15	−1.67*	2.31**	2.98***	1.59#	−0.16	0.88
M2	−4.99***	1.15		−0.91	2.70***	3.55***	3.84***	0.84	−0.20
M3	−3.90***	1.67*	0.91		3.19***	4.01***	4.31***	1.33	−0.97
M4	−6.65***	−2.31**	−2.70***	−3.19***		1.12	−0.23	−1.64#	2.57**
M5	−7.09***	−2.98***	−3.55***	−4.01***	−1.12		−0.97	−2.33**	3.19***
M6	−6.60***	−1.59#	−3.84***	−4.31***	0.23	0.97		−1.44	3.54***
M7	−4.61***	0.16	−0.84	−1.33	1.64#	2.33**	1.44		0.66
M8	−4.27***	−0.88	0.20	0.97	−2.57**	−3.19***	−3.54***	−0.66	

Note: ***, **,* and # denote 1 %, 5 %, 10 %, and 15 % significance levels, respectively. The positive sign of the DM value indicates that F2 is better F1. The negative sign of the DM value indicate that F1 is better F2.

Table 5
The SMAPE of M0, M1, M2, M3, M4, M5, M6, M7, and M8.

Model	M0	M1	M2	M3	M4	M5	M6	M7	M8
SMAPE	10.07	6.25	6.58	7.06	5.29	4.89	5.20	6.14	6.53

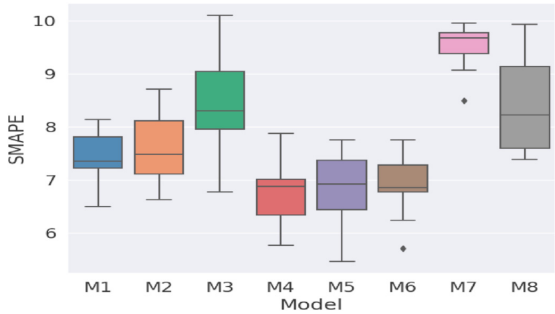


Fig. 10. The SMAPEs of 10 experiments for M1, M2, M3, M4, M5, M6, M7, and M8.

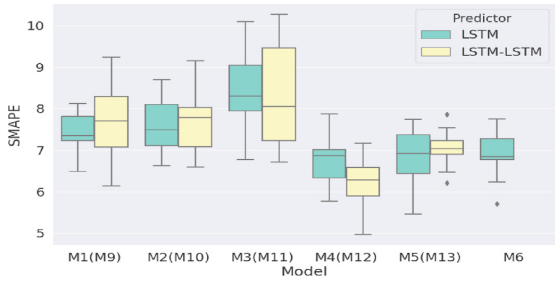


Fig. 11. The comparison of SMAPEs between two-step LSTM models and two-stage LSTM-LSTM models.

Table 6
The results of the one-sided DM test when comparing two-step models (F1) and two-stage models (F2).

F1	M1	M2	M3	M4	M5
	−0.6039	0.2222	2.4556 ***	2.4524 ***	−1.7053 ***
F2	M9	M10	M11	M12	M13

Note: ***, **,* and # denote 1 %, 5 %, 10 %, and 15 % significance levels, respectively. The positive sign: F2 is better F1. The negative sign: F1 is better F2.

the German market has the most electricity cables and the highest electricity exports to the Nordic market, as shown in Figs. 1 and 2. This indicates that it is critical to consider features from cross-border markets with increasing interconnections across Europe for EPF. Besides, electricity prices have more impact on EPF than the features from cross-border electricity trade. From Fig. 15b, we can see that the features have asymmetric predictive influence on electricity price. For instance, the impact of DE consumption (F35) on EPF has a long-tail reaching to the right but not to the left. This indicates that German over-consumption can result in high Nordic electricity prices, but scarce consumption cannot significantly lower the price.

Moreover, we detected the relations between different types of features and the predicted price with their dependence plots. Fig. 16a demonstrates the negative association between DE consumption and its conditional expectation of the predicted price. If the DE consumption is high, then its value tends to revert to its expectation. Thus, the downward expectation of DE consumption will lead to the expected decline of the import demand from the Nordic market, which further decreases the expectation of the Nordic price. Fig. 16b represents the change in predicted price as DE consumption changes. Vertical dispersion at a single value of DE consumption represents the interaction effects with other features. For example, the interaction effect of DE consumption with the Nordic production (F19) is shown in Fig. 16c. The dependence plot highlights that the impact of DE consumption differs with different levels of the Nordic production. The results reveal that the Nordic price is less sensitive to the German power consumption when the Nordic electricity is oversupplied. It indicates, in such a case, that the information from the Nordic market rather than cross-border countries drives the price prediction.

From Fig. 17, we can see that the majority of the EUR/NOK exchange rates (F43) have no contribution to the prediction of the Nordic price (the y-axis value of the dots is zero). In addition, there is no obvious interaction effect of the Nordic productions and the exchange rates on the price. Thus, the predictive importance of exchange rate is extremely limited.

From Fig. 18, we can find that the predicted price is expected to increase when observing a high DK1 → DE cross-border electricity flow, indicating a relatively low current Nordic price. By contrast, a high flow from DE to DK1 implies that the Nordic price is relatively high and expected to decline. In addition, it can be seen via an interaction effect of the DK1 ↔ DE flow with the DE production prognosis that the flow has less impact on the predicted price, with high expected production in Germany. The high production prognosis from cross-border countries will lead to a sharp decline in the expected cross-border transmission. Thus, the impact of the cross-border flow on the Nordic price formation on the following day will decrease significantly.

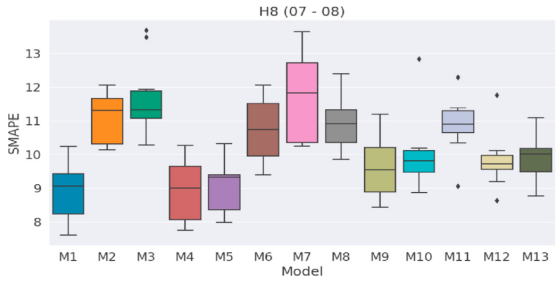


Fig. 12. The SMAPEs of 10 experiments for predicting H8.

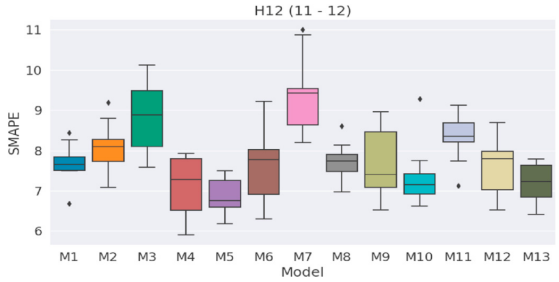


Fig. 13. The SMAPEs of 10 experiments for predicting H12.

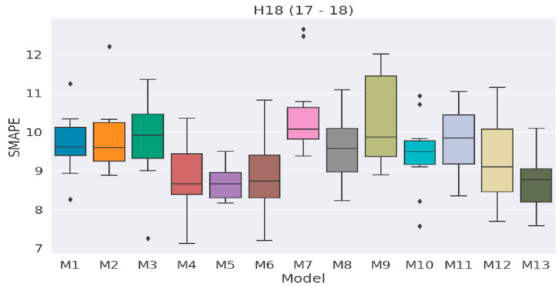


Fig. 14. The SMAPEs of 10 experiments for predicting H18.

Last but not least, not all of the cross-border electricity flows and flow deviations are helpful for forecasting. The reason for this is that, in many cases, the flow capacity is fully occupied. The lack of variability results in their inability to provide useful information for forecasting. An example of the flow and flow deviations between FI and Russia can be seen in Fig. 19. From Fig. 19, it is evident that the majority of flow deviations are zero. The findings indicate that the non-selection of flow deviations from M4 is essential and reasonable. However, the capacity utilisation indicates the potential for more electrical power transmission across the Europe-wide market which would increase the overall socio-economic benefits.

5.3. Discussion of practical importance

The obtained results in the empirical study show that the LSTM-based hybrid models with various features from cross-border markets have considerably accurate prediction results for electricity price. An accurate prediction can be highly beneficial for the electricity market participants in practice. A power market firm that is capable of forecasting the volatile electricity price with a reasonable level of accuracy can reduce trading risk and maximise profits in the day-ahead market by adjusting its bidding strategy and the schedule for production or consumption. More specifically, a 1 % improvement in MAPE of forecast accuracy (within a 5 %–14 % range) leads to about a 0.1–0.35 % cost reduction [82]. On average, a 1 % reduction in the MAPE of short-term price forecasts can result in savings of $1.5 million per year for a typical medium-sized utility company with 5-GW peak load [29,83]. Furthermore, electricity is economically non-storable, and the imbalance between production and consumption can result in power system instability [84]. Accurate electricity forecasting allows energy firms to efficiently organise production or consumption, and this improves the stability of the power system.

In view of the findings from the analysis of feature impact, some implications are important for policy makers to improve cross-border trading in an integrated European power market:

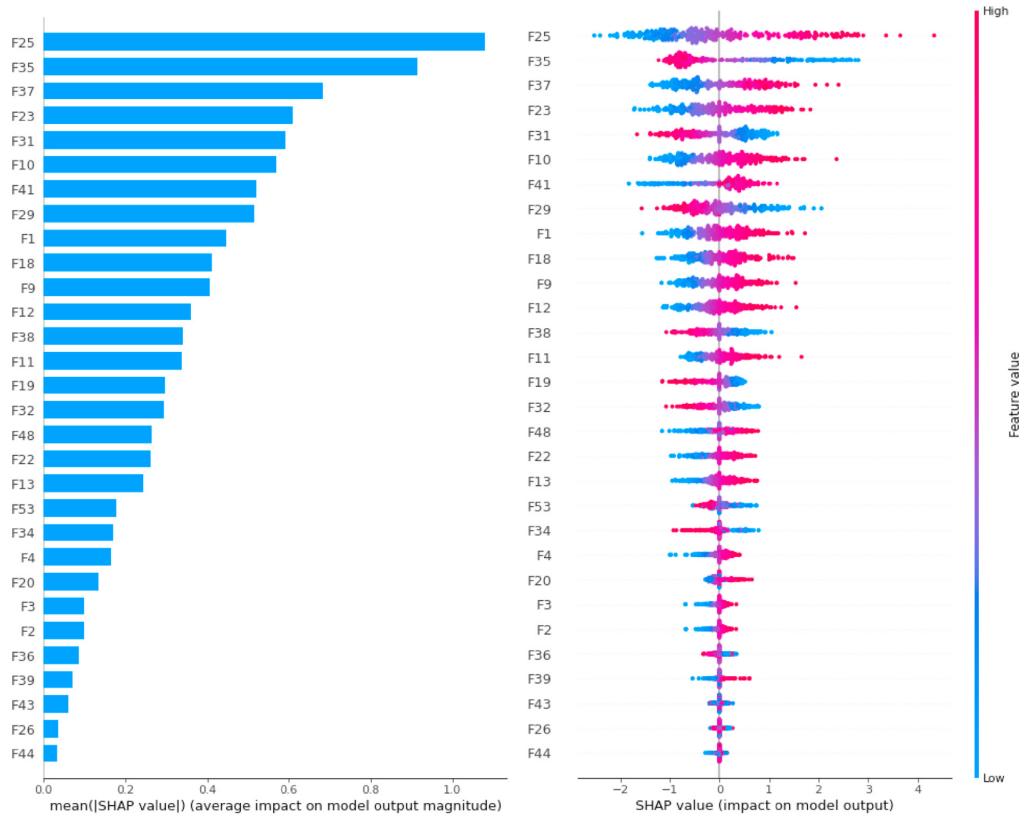
1. The external trading capacities from the German market play a salient role in the generation of Nordic electricity price, and an increasing influence[7] is expected. Thus, all trading capacity between the Nordic and German markets allocated to Nord Pool for implicit auction in the day-ahead price formation could lead to a notable contribution to achieve better allocation of cross-border network capacity, such as the Nordic and Baltic bidding areas.
2. The German production prognosis has a significant predictive impact on the Nordic price. However, the progressive introduction of intermittent renewable energies in Germany[8] [85] makes it difficult to yield accurate predictions. Thus, it is essential to establish formal obligations for cross-border markets to collaborate by sharing useful prospective information. It can better serve the effective demand needed of the electricity trade.
3. The presence of physical transmission constraints can cause a substantial disruption to the market integration and the network congestion implies the shrinkage of the commercial capacity. Thus, an optimal network at a European level should be constructed by a decisive plan.

6. Conclusion

In this paper, we present three LSTM-based hybrid architectures for the EPF. This study puts emphasis on the influence of feature selection methods in the proposed hybrid models. In particular, we compare the prediction performance of the two-step feature selection, the autoencoder, and two-stage feature selection models based on the empirical study on the Nord Pool day-ahead system price. In addition, we employ a SHAP method to evaluate the importance and impact of the features on predicting this price. The main findings are the following: (1) We conclude that the different feature selection methods will lead to different feature selections. As input, diverse features will have a comparably significant impact on the performance of LSTM-based predictive models. (2) Compared to CNN-LSTM and ConvLSTM, LSTM-LSTM is a better autoencoder structure for EPF. (3) The two-stage models can

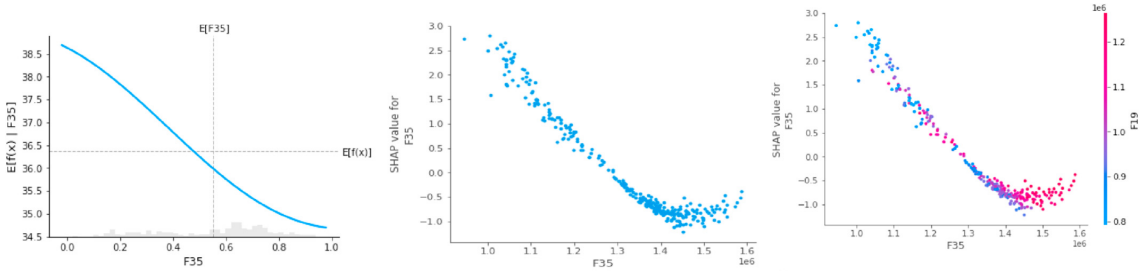
[7] The NordLink, the power cable being built between Norway and Germany, is expected to commence operation in 2021.

[8] Renewable power generation covered more than 46 % of Germany's power consumption in 2020. The forecasting information of intermittent renewable energies, wind and photovoltaic infeed forecasts, in Germany are essential supply-side variables for the adjustment positions of intraday trading and are updated every 15 min until the physical delivery of electricity.



(a) Features ranking **(b)** Features impact

Fig. 15. The feature ranking and feature impact of the selected features of RFE-SVR. (a) Bar chart of the average SHAP value magnitude showing the importance of the features. (b) A set of beeswarm plots, where each dot corresponds to an individual day-ahead price. The dot's position on the x-axis shows the impact that feature has on the model's prediction for that price. Multiple dots landing at the same x position pile up to show density.



(a) SHAP partial dependence plot of DE consumption **(b)** SHAP dependence plot of DE consumption **(c)** SHAP dependence plot of the interaction effect of DE consumption with the Nordic production (F19)

Fig. 16. The SHAP partial dependence (a) and dependence (b, c) plot of DE consumption (F35). $\mathbb{E}[F35]$ is the expectation of the DE consumption, and $\mathbb{E}[f(x)]$ is the expectation of the Nordic price. The grey histogram in (a) shows the distribution of the feature in the test dataset. For (a), the x-axis is the normalised DE consumption. The x-axes in (b) and (c) are the real values of the consumption.

improve the forecasting accuracy of two-step models to some extent. The superior feature selection from the RFE-SVR model allows the autoencoder model to detect more meaningful information for more accurate predictions. (4) The features from the German market (with the most power cables linking to Nord Pool) are more significant for EPF than others. This indicates that more interconnections will increase the cross-border influence on EPF. (5) Compared to other features, the exchange rates are relatively

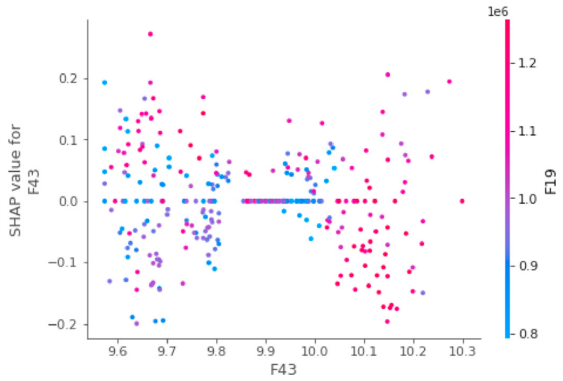


Fig. 17. The SHAP dependence plot of EUR/NOK (F43). The interaction effect of EUR/NOK with the Nordic production (F19).

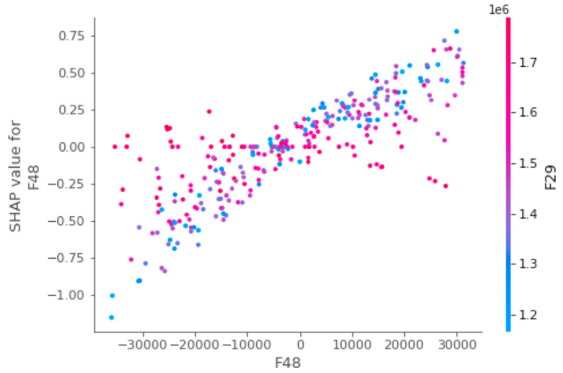


Fig. 18. The SHAP dependence plot of DK1 ↔ DE flow (F48). The interaction effect of DK1 ↔ DE flow with DE production prognosis (F29).

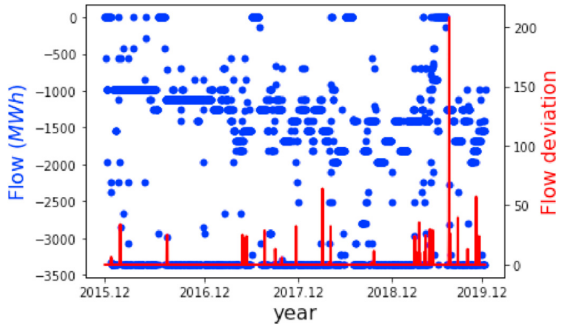


Fig. 19. FI ↔ Russia flow (F54) (blue dots) versus FI ↔ Russia flow deviation (F62) (red line). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

less important. (6) Flow deviation cannot significantly contribute to the price prediction because of its lack of variability. In many cases, the expected flow capacity is fully occupied. The network congestion implies that more interconnections are expected for an efficient Europe-wide electricity market.

For future studies, several extensions of the current study can be developed. Indeed, although the forecasting performance of the proposed models is considerable, we did not conduct an extensive grid search to optimise hyperparameters. It is reasonable to believe that the LSTM-based models with more comprehensive architecture will achieve better forecasting performance. The results will benefit spot electricity traders and policymakers, who make decisions based on accurate price predictions. Moreover, we envision that more testing on other feature selection models can obtain more and different feature selection subsets. They can provide more possibilities for researchers and industries to understand how different features affect prediction accuracy. Finally, the study was carried out using the data from the Nord Pool market, but the generality of the proposed models ensures a possible application to other integrated markets, such as EPEX and OMIE.

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.

Acknowledgement

This work acknowledges research support by COST Action “Fintech and Artificial Intelligence in Finance - Towards a transparent financial industry” (FinAI) CA19130, and has been performed within the +CityxChange[9] (Positive City ExChange) project under the Smart Cities and Communities topic that has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 824260. Critical comments and advice from Florentina Paraschiv, Rüdiger Kiesel, and Frode Kjærland are gratefully acknowledged. The computations were performed on resources provided by UNINETT Sigma2 - the National Infrastructure for High Performance Computing and Data Storage in Norway.

A

A.1. PSO

Each particle has knowledge about its current velocity, its own past best solution ($\vec{p}(t)$), and the current global best solution ($\vec{g}(t)$). Based on this information, each particle's velocity is updated such that it moves closer to the global best and its past best solution at the same time. The velocity update is performed according to the following equation:

$$\vec{v}(t+1) = \omega\vec{v}(t) + c_1 r_1\left(\vec{p}(t) - \vec{x}(t)\right) + c_2 r_2\left(\vec{g}(t) - \vec{x}(t)\right) \tag{A.1}$$

where c_1 and c_2 are constants defined beforehand, which determine the significance of $\vec{p}(t)$ and $\vec{g}(t)$. $\vec{v}(t)$ is the velocity of the particle, $\vec{x}(t)$ is the current particle position, r_1 and r_2 are random numbers from the interval [0,1], and ω is a constant ($0 \leq \omega \leq 1$). The new position is calculated by summing the previous position and the new velocity as follows:

$$\vec{x}(t+1) = \vec{x}(t) + \vec{v}(t+1) \tag{A.2}$$

[9] https://cityxchange.eu/.

This iterative process is repeated until a stopping criterion is satisfied.

A.2. GA-ELM workflow

Fig. A.20 shows the workflow of PSO-ELM and GA-ELM models for feature selection. The process flow of GA-ELM can be described as follows:

Step 1: Initialise the population with a set of random individuals, each individual representing a particular subset of features. For a specific individual (feature set), the features are encoded as "1" or "0", as shown in Fig. A.20. "1" means that the feature is selected, and "0" means that it is not selected. **Step 2**: The selected features are the input for the ELM. The prediction results of the ELM are used to evaluate the fitness value of the individuals. The fitness value is calculated based on the MSE. **Step 3**: Select the best individual with regard to the fitness value. If its fitness is higher than the lowest value in the existing mating pool, it will replace the individual with the worst fitness. Furthermore, the global optimum will be updated accordingly. **Step 4**: The child individuals are generated by crossover and mutation. The new generation is composed of a set of new individuals that are encoded and prepared to be evaluated. The whole process continues until meeting the iteration terminal. The best feature subset in the mating pool is the optimal selection.

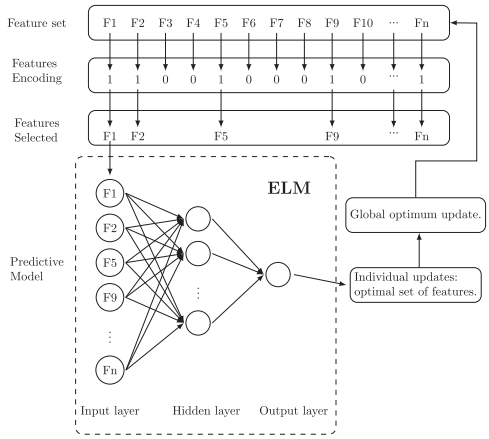


Fig. A.20. The workflow of the two-step wrapper-based feature selection model.

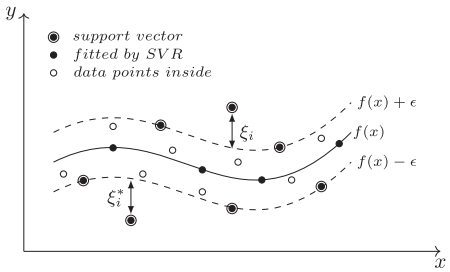


Fig. A.21. Fitted SVR.

A.3. SVR

To minimise the forecasting errors, SVR individualises the hyperplane by maximising the margin. To solve a nonlinear regression problem, the following linear estimation function is considered as follows [86]:

$$f(x) = (\mathbf{w} \times \Phi(x)) + b \tag{A.3}$$

where $\mathbf{w}$ is the parameter vector, $\Phi(x)$ is a kernel function and b is a bias vector. The function formulation of the SVR model can be transformed into the following convex minimisation problem:

$$\min \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{A.4}$$

subject to the following constraints:

$$(\mathbf{w} \times \Phi(x_i) + b) - y_i \le \epsilon + \xi_i$$
$$y_i - (\mathbf{w} \times \Phi(x_i) + b) \le \epsilon + \xi_i^*$$
$$\xi_i, \xi_i^* \ge 0; i = 1, 2, \ldots, n$$

where C is a regularisation constant and ξ_i and ξ_i^* are slack variables, which are used to handle the situation where no such function $f(x)$ exists to satisfy the constraint $|y_i - (\mathbf{w} \times x_i + b)| \le \epsilon$ for all points. They are regarded as the soft margin to allow regression errors, ϵ, to exist up to ξ_i and ξ_i^* and still satisfy the constraint. Only the points outside the ϵ-radius contribute to the final cost. The error parameter, ϵ, represents the region of the tube located around the regression function, $f(x)$, as shown in Fig. A.21.

A.4. The statistical details of the model performance

Table A.7
The MAD (%) results for M1, M2, M3, M4, M5, M6, M7 and M8.

	M1	M2	M3	M4	M5	M6	M7	M8
count	10	10	10	10	10	10	10	10
mean	2.78	2.99	3.31	2.54	2.61	2.67	3.67	3.28
std	0.17	0.29	0.40	0.25	0.30	0.25	0.15	0.38
min	2.42	2.58	2.64	2.16	2.04	2.20	3.30	2.84
25 %	2.72	2.77	3.10	2.34	2.45	2.59	3.65	2.99
50 %	2.77	2.95	3.25	2.59	2.65	2.65	3.69	3.18
75 %	2.92	3.24	3.56	2.63	2.74	2.85	3.73	3.60
max	2.98	3.41	3.93	2.99	3.00	3.09	3.89	3.95

Note: 25 %, 50 %, and 75 % denote 25 %, 50 %, and 75 % percentiles.

Table A.8
The RMSE (%) results for M1, M2, M3, M4, M5, M6, M7 and M8.

	M1	M2	M3	M4	M5	M6	M7	M8
count	10	10	10	10	10	10	10	10
mean	3.60	3.79	4.22	3.25	3.46	3.33	4.74	3.99
std	0.24	0.42	0.55	0.33	0.43	0.32	0.29	0.46
min	3.17	3.15	3.35	2.61	2.79	2.85	4.13	3.50
25 %	3.51	3.45	3.87	3.04	3.27	3.13	4.62	3.60
50 %	3.64	3.78	4.16	3.30	3.50	3.30	4.90	3.92
75 %	3.71	4.05	4.55	3.45	3.71	3.47	4.91	4.33
max	3.96	4.51	5.11	3.73	4.08	3.91	4.97	4.77

Note: 25 %, 50 %, and 75 % denote 25 %, 50 %, and 75 % percentiles.

Table A.9
The MAPE (%) results for M1, M2, M3, M4, M5, M6, M7 and M8.

	M1	M2	M3	M4	M5	M6	M7	M8
count	10	10	10	10	10	10	10	10
mean	7.24	7.83	8.75	6.66	6.79	7.01	9.31	8.73
std	0.46	0.73	1.05	0.65	0.78	0.68	0.47	1.06
min	6.28	6.82	6.89	5.72	5.32	5.67	8.56	7.50
25 %	7.05	7.22	8.19	6.18	6.35	6.84	9.08	7.90
50 %	7.24	7.71	8.62	6.73	6.85	6.95	9.16	8.44
75 %	7.61	8.49	9.35	6.87	7.15	7.45	9.50	9.59
max	7.80	8.93	10.37	7.86	7.91	8.09	10.27	10.66

Note: 25 %, 50 %, and 75 % denote 25 %, 50 %, and 75 % percentiles.

Table A.10
The SMAPE (%) results for M1, M2, M3, M4, M5, M6, M7 and M8.

	M1	M2	M3	M4	M5	M6	M7	M8
count	10	10	10	10	10	10	10	10
mean	7.42	7.58	8.46	6.76	6.85	6.87	9.51	8.38
std	0.50	0.67	0.95	0.65	0.72	0.60	0.44	0.90
min	6.49	6.63	6.77	5.76	5.46	5.71	8.49	7.38
25 %	7.22	7.10	7.95	6.33	6.43	6.77	9.38	7.59
50 %	7.34	7.48	8.30	6.87	6.91	6.85	9.67	8.22
75 %	7.81	8.10	9.05	7.01	7.37	7.27	9.77	9.13
max	8.13	8.70	10.09	7.87	7.74	7.75	9.95	9.93

Note: 25 %, 50 %, and 75 % denote 25 %, 50 %, and 75 % percentiles.

Credit author statement

Wei Li: Conceptualization, Data curation, Validation, Formal analysis, Methodology, Software, Resources, Writing - Original Draft, Writing - Review & Editing. Denis Becker: Conceptualization, Supervision, Validation, Writing - Review & Editing.

References

[1] Weron R. Modeling and forecasting electricity loads and prices: a statistical approach. Wiley; 2006.
[2] Bunn D. Modelling prices in competitive electricity markets. Wiley; 2004.
[3] Nogales FJ, Contreras J, Conejo AJ, Espinola R. Forecasting next-day electricity prices by time series models. IEEE Trans Power Syst 2002;17(2):342–8. https://doi.org/10.1109/TPWRS.2002.1007902.
[4] Bunn DW. Forecasting loads and prices in competitive power markets. Proc IEEE 2000;88(2):163–9. https://doi.org/10.1109/5.823996.
[5] Weron R. Electricity price forecasting: a review of the state-of-the-art with a look into the future. Int J Forecast 2014;30(4):1030–81. https://doi.org/10.1016/j.ijforecast.2014.08.008.
[6] Nowotarski J, Weron R. Recent advances in electricity price forecasting: a review of probabilistic forecasting. Renew Sustain Energy Rev 2018;81:1548–68. https://doi.org/10.1016/j.rser.2017.05.234.
[7] Ventosa M, Baillo A, Ramos A, Rivie M. Electricity market modeling trends. Energy Pol 2005;33(7):897–913. https://doi.org/10.1016/j.enpol.2003.10.013.
[8] Kiose D, Voudouris V. The acewem framework: an integrated agent-based and statistical modelling laboratory for repeated power auctions. Expert Syst Appl 2015;42(5):2731–48. https://doi.org/10.1016/j.eswa.2014.11.024.
[9] Burger M, Schindlmayr G, Graeber B. Managing energy risk: an integrated view on power and other energy markets. Wiley; 2007.
[10] Islyaev S, Date P. Electricity futures price models: calibration and forecasting. Eur J Oper Res 2015;247(1):144–54. https://doi.org/10.1016/j.ejor.2015.05.063.
[11] Weron R, Misiorek A. Forecasting spot electricity prices: a comparison of parametric and semiparametric time series models. Int J Forecast 2008;24(4):744–63. https://doi.org/10.1016/j.ijforecast.2008.08.004.
[12] Conejo AJ, Contreras J, Espínola R, Plazas MA. Forecasting electricity prices for a day-ahead pool-based electric energy market. Int J Forecast 2005;21(3):435–62. https://doi.org/10.1016/j.ijforecast.2004.12.005.
[13] Misiorek A, Trueck S, Weron R. Point and interval forecasting of spot electricity prices: linear vs non-linear time series models. Stud Nonlinear Dynam Econom 2006;10(3). https://doi.org/10.2202/1558-3708.1362.
[14] Gonzalez JP, Roque AMSMS, Pérez EA. Forecasting functional time series with a new Hilbertian ARMAX model: application to electricity price forecasting. IEEE Trans. Power Syst. 2018;33(1):545–56. https://doi.org/10.1109/TPWRS.2017.2700287.
[15] Catalao J, Mariano S, Mendes V, Ferreira L. Short-term electricity prices forecasting in a competitive market: a neural network approach. Elec Power Syst Res 2007;77(10):1297–304. https://doi.org/10.1016/j.epsr.2006.09.022.
[16] Keles D, Scelle J, Paraschiv F, Fichtner W. Extended forecast methods for day-ahead electricity spot prices applying artificial neural networks. Appl Energy 2016;162:218–30. https://doi.org/10.1016/j.apenergy.2015.09.087.
[17] Peter S, Raglend I. Sequential wavelet-ANN with embedded ANN-PSO hybrid electricity price forecasting model for indian energy exchange. Neural Comput Appl 2017;28:2277–92. https://doi.org/10.1007/s00521-015-2141-3.
[18] Hinton G, Deng L, Yu D, Dahl GE, rahman Mohamed A, Jaitly N, et al. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. IEEE Signal Process Mag 2012;29:82–97. https://doi.org/10.1109/MSP.2012.2205597.
[19] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate [cs.CL]. 2014. Available from: 1409.0473. http://arxiv.org/abs/1409.0473.
[20] Li L, Yuan Z, Gao Y. Maximization of energy absorption for a wave energy converter using the deep machine learning. Energy 2018;165:340–9. https://doi.org/10.1016/j.energy.2018.09.093.
[21] Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult. IEEE Trans Neural Network 1994;5(2):157–66. https://doi.org/10.1109/72.279181.
[22] Lago J, Ridder FD, Schutter BD. Forecasting spot electricity prices: deep learning approaches and empirical comparison of traditional algorithms. Appl Energy 2018;221:386–405. https://doi.org/10.1016/j.apenergy.2018.02.069.
[23] Chang Z, Zhang Y, Chen W. Electricity price prediction based on hybrid model of adam optimized LSTM neural network and wavelet transform. Energy 2019;187:115804. https://doi.org/10.1016/j.energy.2019.07.134.
[24] Kuo PH, Huang CJ. An electricity price forecasting model by hybrid structured deep neural networks. Sustainability 2018;10(4). https://doi.org/10.3390/su10041280.
[25] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining, inference and prediction. 2 ed. Springer; 2009.
[26] Li J, Cheng K, Wang S, Morstatter F, Trevino RP, Tang J, et al. Feature selection: a data perspective. ACM Comput Surv 2017;50(6). https://doi.org/10.1145/3136625.
[27] Ziel F, Steinert R, Husmann S. Forecasting day ahead electricity spot prices: the impact of the exaa to other european electricity markets. Energy Econ 2015;51:430–44. https://doi.org/10.1016/j.eneco.2015.08.005.
[28] Panapakidis IP, Dagoumas AS. Day-ahead electricity price forecasting via the application of artificial neural network based models. Appl Energy 2016;172:132–51. https://doi.org/10.1016/j.apenergy.2016.03.089.
[29] Lago J, De Ridder F, Vrancx P, De Schutter B. Forecasting day-ahead electricity prices in europe: the importance of considering market integration. Appl Energy 2018;211:890–903. https://doi.org/10.1016/j.apenergy.2017.11.098.
[30] Uribe JM, Mosquera-López S, Guillen M. Characterizing electricity market integration in nord pool. Energy 2020;208:118368. https://doi.org/10.1016/j.energy.2020.118368.
[31] Marcjasz G, Lago J, Weron R. Neural networks in day-ahead electricity price forecasting: single vs. multiple outputs [stat.AP]. 2020. Available from: 2008.08006. https://arxiv.org/abs/2008.08006.
[32] Johannesen NJ, Kolhe M, Goodwin M. Deregulated electric energy price forecasting in nordpool market using regression techniques. In: In: 2019 IEEE sustainable power and energy conference (iSPEC); 2019. p. 1932–8. https://ieeexplore.ieee.org/abstract/document/8975173.
[33] Chandrashekar G, Sahin F. A survey on feature selection methods. Comput Electr Eng 2014;40(1):16–28. https://doi.org/10.1016/j.compeleceng.2013.11.024. 40th-year commemorative issue.
[34] Sanz H, Valim C, Oller JM, Reverter F. SVM-RFE: selection and visualization of the most relevant features through non-linear kernels. BMC Bioinf 2018;19(1):432. https://doi.org/10.1186/s12859-018-2451-4.
[35] Guyon I, Elisseeff A. An introduction to variable and feature selection. J Mach Learn Res 2003;3:1157–82.
[36] Chen X, Zeng X, van Alphen D. Multi-class feature selection for texture classification. Pattern Recogn Lett 2006;27(14):1685–91. https://doi.org/10.1016/j.patrec.2006.03.013.
[37] Nguyen HB, Xue B, Liu I, Andreae P, Zhang M. New mechanism for archive maintenance in PSO-based multi-objective feature selection. Soft Comput. 2016;20:3927–46. https://doi.org/10.1007/s00500-016-2128-8.
[38] Shang L, Zhou Z, Liu X. Particle swarm optimization-based feature selection in sentiment classification. Soft Comput. 2016;20:3821–34. https://doi.org/10.1007/s00500-016-2128-8.
[39] Zhou Y, Zhou N, Gong L, Jiang M. Prediction of photovoltaic power output based on similar day analysis, genetic algorithm and extreme learning machine. Energy 2020;204:117894. https://doi.org/10.1016/j.energy.2020.117894.
[40] Krishnan G S, S S. A novel GA-ELM model for patient-specific mortality prediction over large-scale lab event data. Appl. Soft Comput. J. 2019;80:525–33. https://doi.org/10.1016/j.asoc.2019.04.019.
[41] Luo P, Zhu S, Han L, Chen Q. Short-term photovoltaic generation forecasting based on similar day selection and extreme learning machine. In: IEEE power and energy society general meeting 2018; 2018. https://doi.org/10.1109/PESGM.2017.8273776. January:1–5.
[42] Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. Mach Learn 2002;46(1):389–422. https://doi.org/10.1023/A:1012487302797.
[43] Tibshirani R. Regression shrinkage and selection via the lasso. J Roy Stat Soc Soc B (Methodological) 1996;58(1):267–88. https://doi.org/10.1111/j:2517-6161:1996:tb02080:x.
[44] Greenfish. Shaping our electrical future: moving towards an integrated european network. 2019. https://www.greenfish.eu/shaping-our-electrical-future-moving-towards-an-integrated-european-network/. [Accessed 13 September 2020].
[45] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;74(8):1735–80. https://doi.org/10.1162/neco.1997.9.8.1735.
[46] Gers FA, Schmidhuber J. Recurrent nets that time and count. In: Proceedings of the IEEE-INNS-ENNS international joint conference on neural networks. IJCNN 2000. Neural computing: new challenges and perspectives for the new millennium, vol. 3; 2000. p. 189–94. https://doi.org/10.1109/IJCNN.2000.861302.
[47] Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Network 2005;18(5):602–10. https://doi.org/10.1016/j.neunet.2005.06.042. IJCNN 2005.
[48] Cho K, van Merrienboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H,

et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation [cs.CL]. 2014. Available from: 1406.1078. https://arxiv.org/abs/1406.1078.
[49] Graves A. Generating sequences with recurrent neural networks [cs.NE]. 2013. Available from: arXiv:1308.0850. https://arxiv.org/abs/1308.0850.
[50] Huang GB, Zhu QY, Siew CK. Extreme learning machine: theory and applications. Neurocomputing 2006;70(1):489–501. https://doi.org/10.1016/j.neucom.2005.12.126.
[51] Saraswathi S, Sundaram S, Sundararajan N, Zimmermann M, Nilsen-Hamilton M. ICGA-PSO-ELM approach for accurate multiclass cancer classification resulting in reduced gene sets in which genes encoding secreted proteins are highly represented. IEEE ACM Trans Comput Biol Bioinf 2011;8(2):452–63. https://doi.org/10.1109/TCBB.2010.13.
[52] Chyzhyk D, Savio A, Graña M. Evolutionary ELM wrapper feature selection for Alzheimer's disease CAD on anatomical brain MRI. Neurocomputing 2014;128:73–80. 10.1016/j.neucom.2013.01.065.
[53] Ahila R, Sadasivam V, Manimala K. An integrated PSO for parameter determination and feature selection of ELM and its application in classification of power system disturbances. Appl Soft Comput 2015;32:23–37. https://doi.org/10.1016/j.asoc.2015.03.036.
[54] Zhang Y, Wang S, Ji G. A comprehensive survey on particle swarm optimization algorithm and its applications. Math Probl Eng 2015;2015:931256. https://doi.org/10.1155/2015/931256.
[55] Whitley D. A genetic algorithm tutorial. Stat Comput 1994;4(2):65–85. https://doi.org/10.1007/BF00175354.
[56] Drucker H, Burges CJC, Kaufman L, Smola A, Vapnik V. Support vector regression machines. In: Mozer MC, Jordan M, Petsche T, editors. Advances in neural information processing systems. vol. 9. MIT Press; 1997. https://proceedings.neurips.cc/paper/1996/file/d38901788c533e8286cb6400b40b386d-Paper.pdf.
[57] Szegedy C, Liu Wei, Jia Yangqing, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. In: 2015 IEEE conference on computer vision and pattern recognition (CVPR); 2015. p. 1–9. https://doi.org/10.1109/CVPR.2015.7298594.
[58] Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks. In: Ghahramani Z, Welling M, Cortes C, Lawrence ND, Weinberger KQ, editors. Advances in neural information processing systems 27. Curran Associates, Inc; 2014. p. 3104–12. https://arxiv.org/abs/1409.3215.
[59] Shi X, Chen Z, Wang H, Yeung DY, Wk Wong, Wc WOO. Convolutional LSTM network: a machine learning approach for precipitation nowcasting. In: Cortes C, Lawrence ND, Lee DD, Sugiyama M, Garnett R, editors. Advances in neural information processing systems 28. Curran Associates, Inc; 2015. p. 802–10. https://arxiv.org/pdf/1506.04214.pdf.
[60] Diebold FX, Mariano RS. Comparing predictive accuracy. J Bus Econ Stat 2002;20(1):134–44. https://doi.org/10.1198/073500102753410444.
[61] Harvey D, Leybourne S, Newbold P. Testing the equality of prediction mean squared errors. Int J Forecast 1997;13(2):281–91. https://doi.org/10.1016/S0169-2070(96)00719-4.
[62] Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. BMC Bioinf 2006;7:91. https://doi.org/10.1186/1471-2105-7-91.
[63] Goodfellow I, Bengio Y, Courville A. Deep learning. The MIT Press; 2016.
[64] McHugh C, Coleman S, Kerr D, McGlynn D. Daily energy price forecasting using a polynomial narmax model. In: Lotfi A, Bouchachia H, Gegov A, Langensiepen C, McGinnity M, editors. Advances in computational intelligence systems. Cham: Springer International Publishing; 2019, ISBN 978-3-319-97982-3. p. 71–82.
[65] Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable ai for trees. Nat. Mach. Intell. 2020;2(1):56–67. https://doi.org/10.1038/s42256-019-0138-9.
[66] Janzing D, Minorics L, Blöbaum P. Feature relevance quantification in explainable ai: a causal problem. 2019. Available from: 1910.13413. https://arxiv.org/abs/1910.13413.
[67] Sundararajan M, Najmi A. The many shapley values for model explanation. 2020. Available from: 1908.08474. https://arxiv.org/abs/1908.08474.
[68] Lundberg S, Lee SI. A unified approach to interpreting model predictions. 2017. Available from: 1705.07874. https://arxiv.org/abs/1705.07874.
[69] Peng Hanchuan, Long Fuhui, Ding C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell 2005;27(8):1226–38. https://doi.org/10.1109/TPAMI.2005.159.
[70] Yu L, Liu H. Efficient feature selection via analysis of relevance and redundancy. J Mach Learn Res 2004;5:1205–24.
[71] Langley P. Selection of relevant features in machine learning. In: In Proceedings of the AAAI Fall symposium on relevance. AAAI Press; 1994. p. 140–4.
[72] Jamian JJ, Abdullah MN, Mokhlis H, Mustafa MW, Bakar AHA. Global particle swarm optimization for high dimension numerical functions analysis. J Appl Math 2014. https://doi.org/10.1155/2014/329193.
[73] Somu N, Raman MRG, Ramamritham K. A deep learning framework for building energy consumption forecast. Renew Sustain Energy Rev 2021;110591:137. https://doi.org/10.1016/j.rser.2020.110591. https://www.sciencedirect.com/science/article/pii/S1364032120308753.
[74] Shao Z, Yang S, Gao F, Zhou K, Lin P. A new electricity price prediction strategy using mutual information-based svm-rfe classification. Renew Sustain Energy Rev 2017;70:330–41. https://doi.org/10.1016/j.rser.2016.11.155. https://www.sciencedirect.com/science/article/pii/S1364032116309297.
[75] Radovic M, Ghalwash M, Filipovic N, Obradovic Z. Minimum redundancy maximum relevance feature selection approach for temporal gene expression data. BMC Bioinf 2017;18(1):9. https://doi.org/10.1186/s12859-016-1423-9.
[76] Kim TY, Cho SB. Predicting the household power consumption using cnn-lstm hybrid networks. In: Yin H, Camacho D, Novais P, Tallón-Ballesteros AJ, editors. Intelligent data engineering and automated learning – ideal 2018. Cham: Springer International Publishing; 2018. p. 481–90.
[77] Kim TY, Cho SB. Predicting residential energy consumption using cnn-lstm neural networks. Energy 2019;182:72–81. https://doi.org/10.1016/j.energy.2019.05.230.
[78] Sultana T, Khan ZA, Javaid N, Aimal S, Fatima A, Shabbir S. Data analytics for load and price forecasting via enhanced support vector regression. In: Barolli L, Xhafa F, Khan ZA, Odhabi H, editors. Advances in internet, data and web technologies. Cham: Springer International Publishing; 2019. p. 259–70.
[79] Nawaz M, Javaid N, Mangla FU, Munir M, Ihsan F, Javaid A, et al. An approximate forecasting of electricity load and price of a smart home using nearest neighbor. In: Barolli L, Hussain FK, Ikeda M, editors. Complex, intelligent, and software intensive systems. Cham: Springer International Publishing; 2020. p. 521–33.
[80] Brusaferri A, Fagiano L, Matteucci M, Vitali A. Day ahead electricity price forecast by NARX model with LASSO based features selection. In: IEEE international conference on industrial informatics (INDIN) 2019; 2019-July. p. 1051–6. https://doi.org/10.1109/INDIN41052.2019.8972263.
[81] Leerbeck K, Bacher P, Junker R, Goranović G, Corradi O, Ebrahimy R, et al. Short-term forecasting of CO2 emission intensity in power grids by machine learning. Appl Energy 2020;277. https://doi.org/10.1016/j.apenergy.2020.115527.
[82] Zareipour H, Canizares CA, Bhattacharya K. Economic impact of electricity market price forecasting errors: a demand-side analysis. IEEE Trans. Power Syst. 2010;25(1):254–62. https://doi.org/10.1109/TPWRS.2009.2030380.
[83] Uniejewski B, Nowotarski J, Weron R. Automated variable selection and shrinkage for day-ahead electricity price forecasting. Energies 2016;9(8). https://doi.org/10.3390/en9080621.
[84] Kaminski V. Energy markets. Risk book. 2013.
[85] Li W, Paraschiv F. Modelling the evolution of wind and solar power infeed forecasts. J Commod Mark 2021:100189. https://doi.org/10.1016/j.jcomm.2021.100189.
[86] Herceg S, Željka Ujević Andrijić, Bolf N. Development of soft sensors for isomerization process based on support vector machine regression and dynamic polynomial models. Chem Eng Res Des 2019;149:95–103. https://doi.org/10.1016/j.cherd.2019.06.034.

Article C

Title:
A Data-driven Explainable Case-based Reasoning Approach for Financial Risk Detection

Authors:
Wei Li
Florentina Paraschiv
Georgios Sermpinis

Reference:
Li W, Paraschiv F, Sermpinis G. A Data-driven Explainable Case-based Reasoning Approach for Financial Risk Detection. Working paper. Submitted to Quantitative Finance.

ISBN 978-82-326-6326-2 (printed ver.)
ISBN 978-82-326-5761-2 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

NTNU

Norwegian University of
Science and Technology