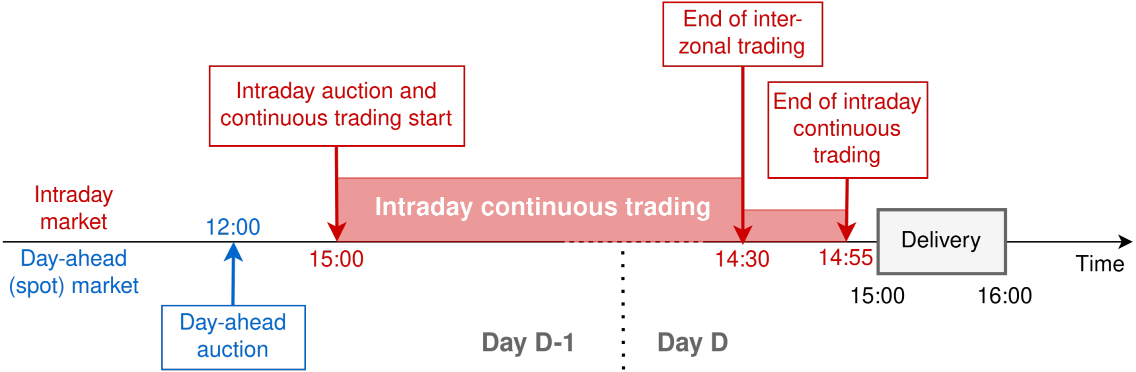
Deep Learning in point, probabilistic and trajectory electricity price forecasting

Grzegorz Marcjasz

supervisor: prof. dr hab. Rafał Weron

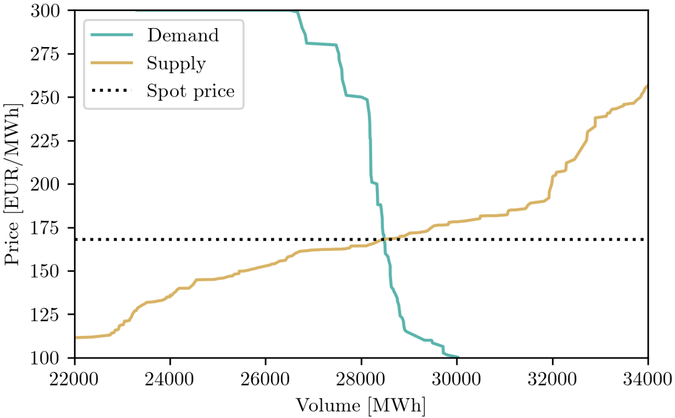
Department of Operations Research and Business Intelligence
Faculty of Management

Day-ahead and intraday electricity trading

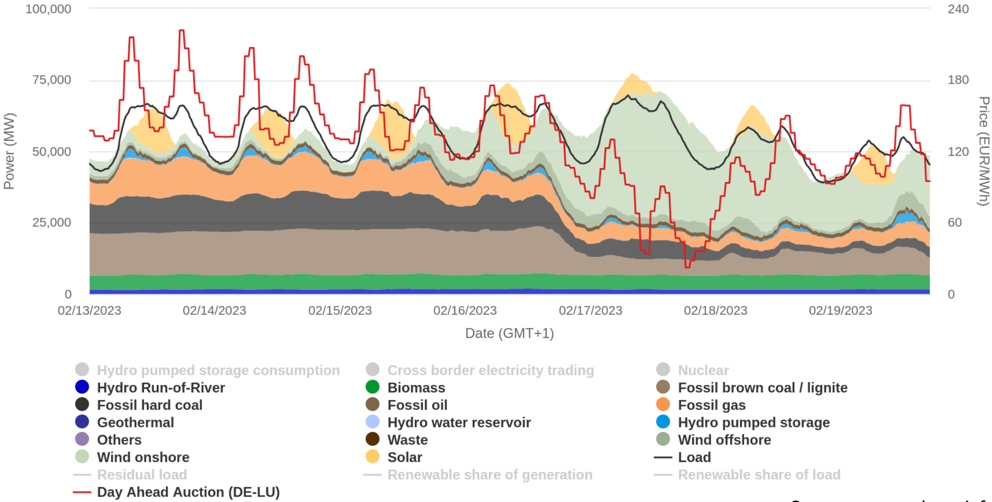


The day-ahead market for electricity

- A two-sided uniform-price auction
- Separate for all hours

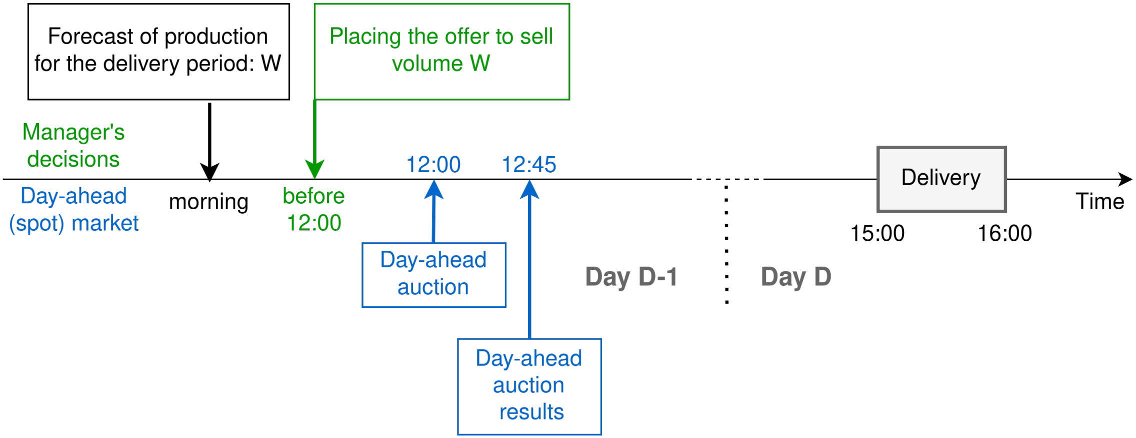


The generation mix and electricity prices

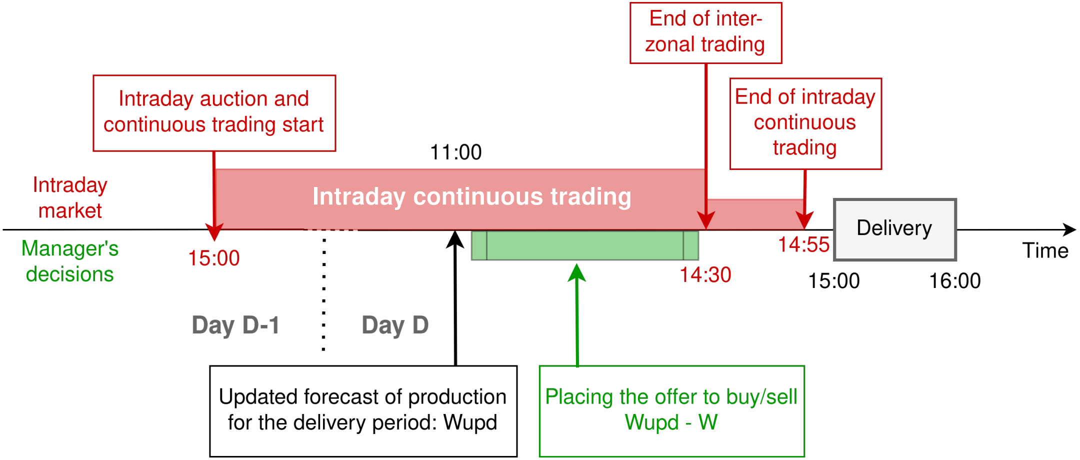


Source: energy-charts.info

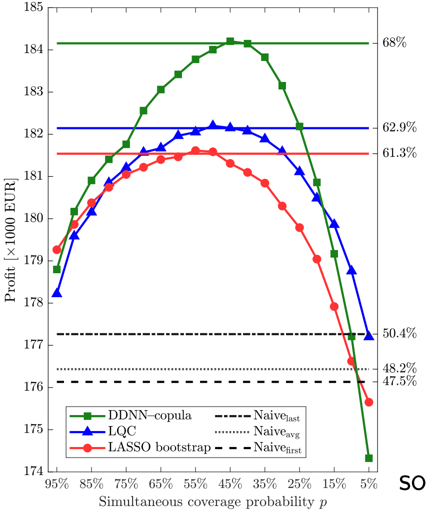
Perspective of a wind power plant manager



Perspective of a wind power plant manager



Is it worth it?



- A small producer in Germany
- 1MWh of imbalance sold on the intraday market
- Trading in the last 3 hours before the delivery
- Dashed lines: results of naive strategies
- Percentages (right scale) of maximum possible profit

source: [P5]

Aim and objectives

Aim: Develop robust, reliable and interpretable DNN-based approaches for short-term point, probabilistic and ensemble forecasting of electricity prices

1. Identify the most common problems encountered in EPF ML research, present a set of best practices and publish open access codes for well-performing benchmark models

energies MDPI

[P1]

Article

Forecasting Electricity Prices Using Deep Neural Networks: A Robust Hyper-Parameter Selection Scheme

Grzegorz Marcjasz

Department of Operations Research and Business Intelligence, Wrocław University of Science and Technology, 50-370 Wrocław, Poland; grzegorz.marcjasz@pwr.edu.pl

Received: 22 July 2020; Accepted: 2 September 2020; Published: 4 September 2020

check for updates

Abstract: Deep neural networks are rapidly gaining popularity. However, their application requires setting multiple hyper-parameters, and the performance relies strongly on this choice. We address this issue and propose a robust ex-ante hyper-parameter selection procedure for the day-ahead

Applied Energy 293 (2021) 116983

Contents lists available at ScienceDirect

Applied Energy

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

ELSEVIER

[P2]

Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark

Jesus Lago [a,*], Grzegorz Marcjasz [b], Bart De Schutter [a], Rafał Weron [b]

[a] Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands
[b] Department of Operations Research and Business Intelligence, Wrocław University of Science and Technology, Wrocław, Poland

ARTICLE INFO

Keywords:
Electricity price forecasting
Regression model
Deep learning

ABSTRACT

While the field of electricity price forecasting has benefited from plenty of contributions in the last two decades, it arguably lacks a rigorous approach to evaluating new predictive algorithms. The latter are often compared using unique, not publicly available datasets and across too short and limited to one market test samples. The proposed new methods are rarely benchmarked against well established and well performing simpler

Aim and objectives

2. Develop an interpretable DNN model for point EPF that outperforms state-of-the-art benchmarks

International Journal of Forecasting 39 (2023) 884–900

Contents lists available at ScienceDirect

International Journal of Forecasting

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

ELSEVIER

[P3]

Neural basis expansion analysis with exogenous variables: Forecasting electricity prices with NBEATSx

Check for updates

Kin G. Olivares [a,*], Cristian Challu [a], Grzegorz Marcjasz [b], Rafał Weron [b], Artur Dubrawski [a]

[a] *Auton Lab, School of Computer Science, Carnegie Mellon University, United States*
[b] *Department of Operations Research and Business Intelligence, Wroclaw University of Science and Technology, Poland*

Aim and objectives

3. Construct distributional DNNs that directly yield predictive distributions and are superior to state-of-the-art probabilistic models in terms of both statistical and economic measures
4. Develop a decision support method that uses distributional DNNs to generate trajectories of ID prices, then use it to construct profitable trading strategies

Energy Economics 125 (2023) 106843

Contents lists available at ScienceDirect

Energy Economics

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

ELSEVIER

[P4]

Distributional neural networks for electricity price forecasting

Grzegorz Marcjasz [a,*], Michał Narajewski [b], Rafał Weron [a], Florian Ziel [b]

[a] Department of Operations Research and Business Intelligence, Wrocław University of Science and Technology, 50-370 Wrocław, Poland

[b] House of Energy Markets and Finance, University of Duisburg-Essen, 45141 Essen, Germany

ARTICLE INFO

JEL classification:
C44
C45
C46
C22
C53
O47

ABSTRACT

We present a novel approach to probabilistic electricity price forecasting which utilizes distributional neural networks. The model structure is based on a deep neural network containing a so-called probability layer, i.e., the outputs of the network are parameters of the normal or Johnson's SU distribution. To validate our approach, we conduct a comprehensive forecasting study complemented by a realistic trading simulation with day-ahead electricity prices in the German market. The proposed distributional deep neural network outperforms state-of-the-art benchmarks by over 7% in terms of the continuous ranked probability score and by 8% in terms

Trading on short-term path forecasts of intraday electricity prices. Part II – Distributional Deep Neural Networks

[P5]

Grzegorz Marcjasz[a,1], Tomasz Serafin[a], Rafał Weron[a]

[a] Department of Operations Research and Business Intelligence, Wrocław University of Science and Technology, 50-370 Wrocław, Poland

Abstract

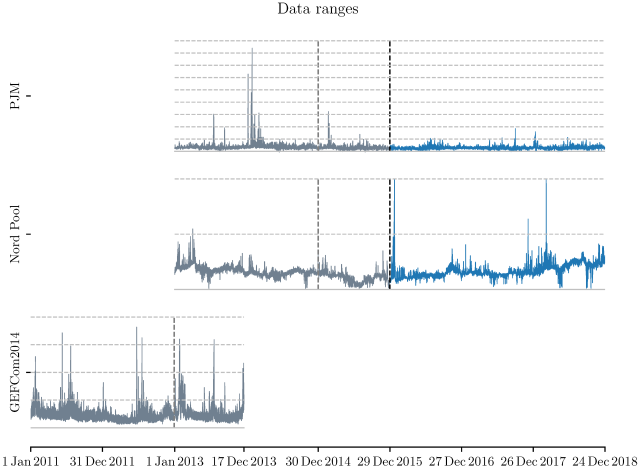
We propose a novel electricity price forecasting model tailored to intraday markets with continuous trading. It is based on distributional deep neural networks with Johnson SU distributed outputs. To demonstrate its usefulness, we introduce a realistic trading strategy for the economic evaluation of ensemble forecasts. Our approach takes into account forecast errors in wind generation for four German TSOs and uses the intraday market to resolve imbalances remaining after day-ahead bidding. We argue that the economic evaluation is crucial and provide evidence that the better performing methods in terms of statistical error metrics do not necessarily lead to higher trading profits.

Keywords: Intraday electricity market, Probabilistic forecast, Path forecast, Prediction bands, Trading strategy, Neural networks

O1: Problems, best practices and benchmarks

[P1] contributions:

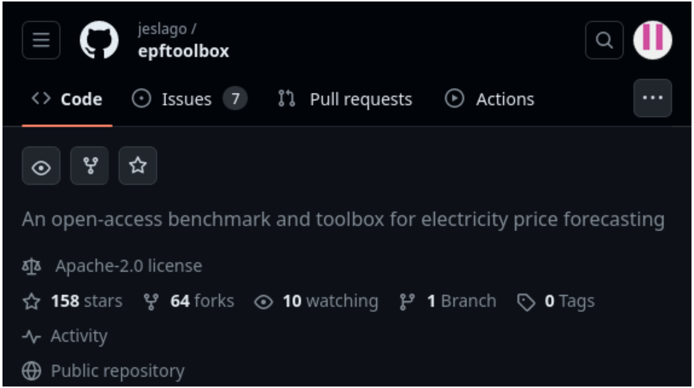
- using data from different markets (than the modeled one) for hyper-parameter optimization
- a robust hyper-parameter selection and aggregation scheme



O1: Problems, best practices and benchmarks

[P2] contributions:

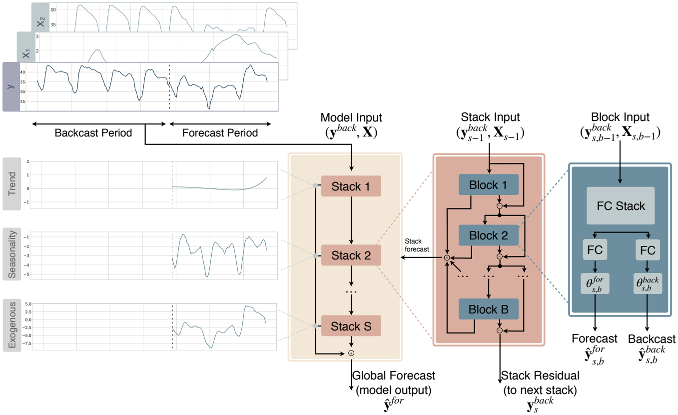
- a comprehensive review of machine learning EPF methods
- identifies shortcomings in the literature
- proposes a set of guidelines
- provides an open-access DNN benchmark



O2: Interpretable DNN model for point EPF

[P3] contributions:

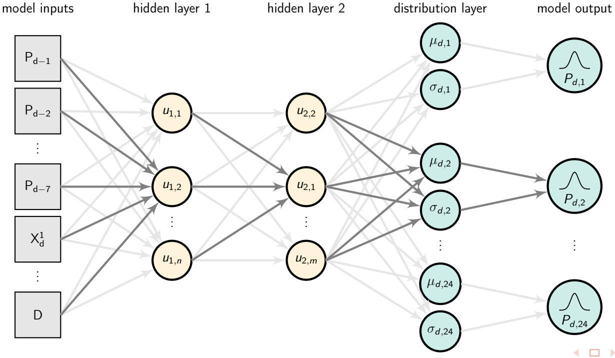
- a novel extension of the NBEATS model: exogenous inputs
- a well-performing model that can be decomposed into seasonal, trend and exogenous factors



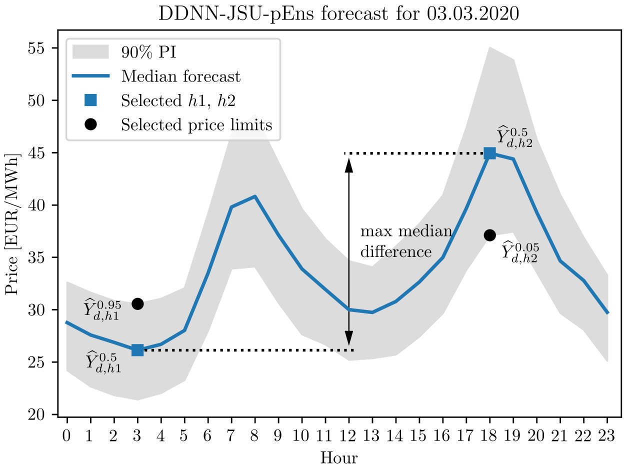
O3: Distributional DNN and economic evaluation

[P4] contributions:

- a DNN that outputs parameters of a distribution
- tested on normal and a Johnson's SU distributions
- forecast accuracy improvement over point NN with QRA



O3: Distributional DNN and economic evaluation



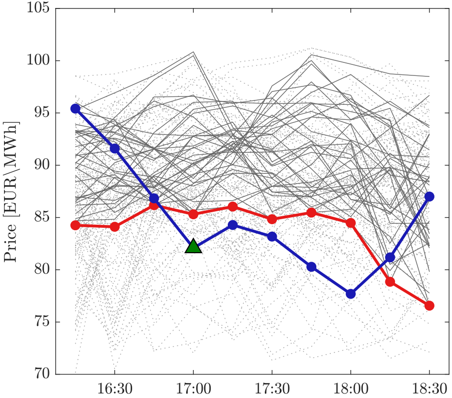
[P4] contributions:

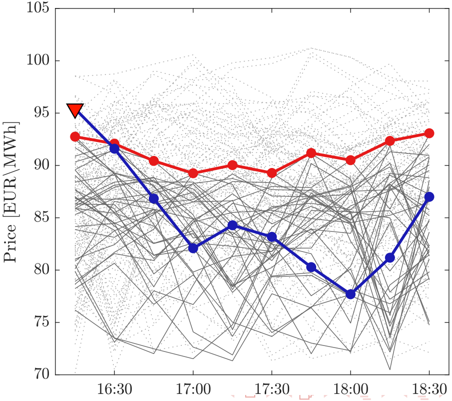
- a real-world market simulation that optimizes battery operation using probabilistic forecasts

O4: Decision support with DDNN-based path forecasts

[P5] contributions:

- an intraday market framework based on the DDNN **[P4]**
- use of path forecasts in automated trading

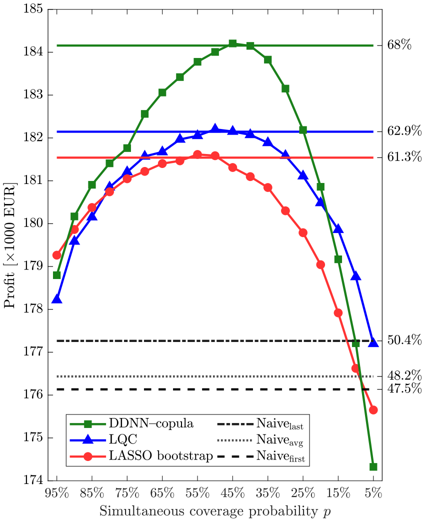




O4: Decision support with DDNN-based path forecasts

[P5] contributions:

- a strategy asssuming the role of wind power plant manager
- outperformance of benchmark models across all test cases
- ca. 5 pp. higher FRTP compared to the LEAR-QR-based model



Key findings

- NN-based models can outperform regression-based approaches
- Forecast averaging is no longer an option, it is a necessity
- NNs are flexible – they can model distributions efficiently
- Hyper-parameter optimization is a crucial step, but automated methods work well
- Probabilistic and trajectory forecasts are useful tools in decision-making

Replies to reviews

1. Justification of the methods chosen

The criteria for choosing a method:

- thorough examination (w.r.t. the hyperparameters, inputs, results stability)
- comparability with the chosen benchmarks (more on that later)
- preference for incrementally added complexity

2. Selection of input features

Choice of the input space based on exploratory studies:

GEFCom2014

$\beta_{h,i}$	Hour 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	600	621	314	116	126	522	623	495	135	316	402	498	490	452	489	456	469	297	411	343	312	452	562	545
2	610	623	471	390	191	352	44	21	11	2	102	78	26	52	13	0	0	0	44	30	112	133	53	51
3	38	453	580	465	271	164	206	75	106	69	155	158	14	191	154	267	298	171	76	164	21	0	0	0
4	205	197	390	550	369	184	2	50	10	0	0	0	0	0	0	9	18	0	0	1	0	0	7	5
5	289	239	396	460	621	282	101	180	41	201	293	154	81	17	1	38	0	62	210	0	128	26	6	0
6	307	467	596	618	604	623	566	365	370	184	184	115	284	316	336	337	257	336	138	147	176	51	233	452
7	8	80	205	158	329	197	623	623	595	613	399	264	80	16	40	15	51	216	317	507	286	326	311	130
8	119	85	164	205	233	275	406	623	619	535	422	151	41	58	38	3	90	204	181	103	251	98	19	5
9	87	0	0	0	6	91	43	67	318	314	73	61	61	56	28	131	30	0	0	0	43	0	14	45
10	269	3	3	0	2	9	261	112	100	170	58	15	64	170	251	304	315	296	268	280	266	278	277	177
11	17	16	17	0	0	0	73	64	242	426	501	556	245	139	7	14	7	123	150	87	1	3	11	0
12	12	17	10	0	0	4	0	0	0	0	177	284	263	168	187	162	74	18	0	15	1	0	2	23
13	61	88	0	78	9	0	0	13	0	117	269	305	324	304	158	93	45	54	53	59	101	83	53	62
14	13	64	0	0	0	14	35	106	5	0	18	51	206	239	182	109	90	5	78	160	183	122	334	19
15	16	0	26	131	244	132	139	198	70	27	40	75	128	280	346	344	356	338	339	121	152	144	20	0
16	347	372	335	443	340	328	234	216	88	26	0	2	119	112	144	206	154	599	188	136	92	64	106	178
17	32	54	20	36	76	468	580	528	566	447	298	212	202	145	257	213	360	121	311	313	344	361	138	146
18	352	367	268	253	143	129	1	132	71	130	104	30	48	63	322	439	488	623	329	16	66	37	211	186
19	354	475	565	477	421	325	385	576	325	365	360	470	491	575	547	541	323	495	623	369	173	358	416	283
20	118	239	169	85	72	387	623	452	298	298	362	388	511	455	538	457	407	258	303	623	332	209	362	272
21	615	561	326	189	247	381	283	623	623	623	608	623	623	623	623	623	623	618	479	556	623	623	368	220
22	326	324	426	385	387	329	353	600	623	623	623	623	623	623	623	623	623	623	623	614	619	623	623	623
23	177	93	227	246	279	238	510	555	574	445	436	340	241	242	265	414	304	359	585	320	372	400	623	571
24	623	623	623	623	623	623	579	594	623	623	623	577	536	407	348	324	279	501	490	493	596	618	623	623

Nord Pool

$\beta_{h,i}$	Hour 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	492	393	338	284	355	146	295	420	385	358	346	351	354	304	451	392	321	388	397	362	407	348	381	594
2	153	115	79	139	100	114	15	10	45	41	39	30	34	28	60	72	46	93	19	12	20	81	35	129
3	190	363	384	348	313	366	4	48	66	6	56	35	39	64	53	27	103	125	71	141	238	285	223	497
4	19	57	73	40	39	0	17	73	110	80	89	22	7	8	6	6	9	0	5	24	21	54	16	36
5	315	260	384	411	461	409	387	305	230	206	140	74	43	132	127	148	112	95	63	18	138	189	165	340
6	323	364	359	484	495	365	421	311	194	250	347	345	396	486	344	381	262	376	335	210	319	299	309	442
7	85	442	438	383	445	550	282	273	270	272	256	350	301	229	335	342	380	386	243	116	131	190	32	162
8	11	0	0	24	17	298	2	300	254	220	95	32	6	61	35	17	91	56	51	27	38	171	239	245
9	129	176	165	135	207	437	251	113	171	298	179	82	102	124	175	247	173	100	308	381	423	381	395	252
10	177	45	56	81	56	109	160	105	141	121	61	144	226	220	259	244	329	333	212	194	158	154	173	7
11	14	57	44	44	191	41	2	357	428	384	355	294	83	141	0	31	146	183	176	78	3	5	7	0
12	228	244	259	314	325	95	282	81	50	6	55	121	251	279	269	195	11	21	4	73	59	28	0	0
13	36	163	158	155	144	12	10	25	20	0	23	87	76	68	92	102	99	42	38	82	105	15	0	0
14	143	33	2	15	32	8	57	20	56	17	30	26	113	110	121	65	31	50	9	0	0	0	13	17
15	157	185	171	153	92	1	16	162	266	193	107	6	1	4	1	13	8	110	4	188	37	4	31	55
16	73	73	72	141	266	208	143	218	322	251	216	125	0	19	31	3	151	329	398	248	2	0	5	84
17	189	319	357	402	287	93	306	370	174	240	459	431	443	436	414	419	365	161	142	189	183	90	216	440
18	214	176	211	369	390	483	442	471	404	489	429	415	436	414	422	411	554	639	659	411	275	199	238	307
19	244	277	295	400	458	212	119	38	110	165	148	54	51	48	103	122	48	104	109	121	50	91	60	38
20	194	275	310	429	456	44	70	180	231	257	226	191	233	196	202	216	252	231	275	260	35	27	31	6
21	88	126	163	162	166	179	339	112	65	55	51	68	64	82	70	72	53	17	48	391	459	217	33	7
22	194	437	625	695	696	654	651	728	728	728	728	728	728	724	721	728	707	657	682	677	683	692	674	375
23	15	86	218	359	279	300	721	553	217	128	341	391	533	553	596	432	360	405	408	374	291	247	656	600
24	728	728	728	728	728	728	542	204	127	179	252	216	173	142	117	154	130	129	188	126	157	105	452	728

Reference: B. Uniejewski, J. Nowotarski and R. Weron (2016) *Automated Variable Selection and Shrinkage for Day-Ahead Electricity Price Forecasting*, Energies 9 (621)

3. Dismissal of cross-border factors

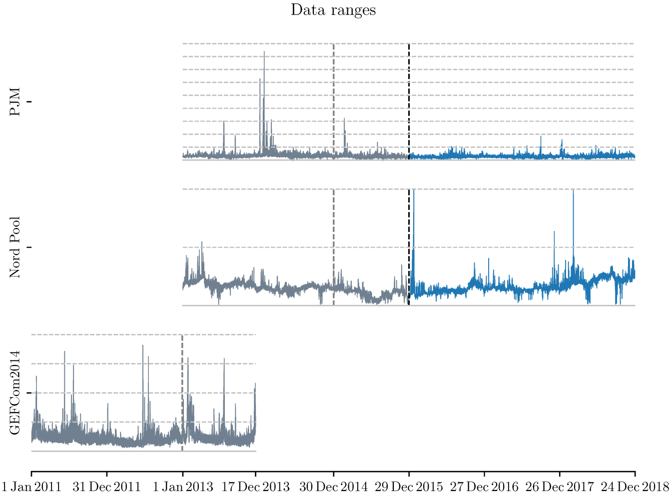
The cross-border factors:

- growing importance in extreme (e.g., high wind) events
- but less important under normal conditions (increasing integration of European markets)
- very complex to incorporate into the model correctly (flow-based market coupling)

[P1] Optimal hyperparameter sets

- NP and PJM: always sigmoid activation, GEFCom: elu chosen as well
- NP: preference of Adam, GEFCom: mixed (AdaGrad, RMSProp, Adamax), PJM: Adagrad and NAdam
- Most of the runs: 500 epochs (2 of the best runs were 200 epochs)
- NP: preference for larger batches

[P1] Validation setting for NN and Lasso models



[P1] Mitigating the risk of fitting parameters that work well only on the validation sample

- Choose the datasets carefully, look for structural changes, filter or transform the data if needed
- Use model ensembling (e.g. train on portions of the data)
- Check the result variance in consecutive runs

[P2] The *state-of-the-art* and the focus of the study

Study built upon the results from 2018 Applied Energy article[1]:

- DNN was the best of 20+ models tested
- lasso (and elastic net) were the best statistical models
- the proposed LEAR model offers a good performance complexity tradeoff
- the literature focuses on the NN-based approaches

Model	sMAPE [%]	Class
DNN	12.34	ML
GRU	13.04	
LSTM	13.06	
MLP	13.27	
SVR	13.29	
SOM-SVR	13.36	
SVR-ARIMA	13.39	
XGB	13.74	
fARX-EN	13.76	SM
CNN	13.91	ML
fARX-Lasso	13.92	SM
RBF	14.77	ML
fARX	14.79	ST
RF	15.39	ML
IHMARX	16.72	ST
DR	16.99	

[1] J. Lago, F. De Ridder and B. De Schutter (2018) *Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms*, Applied Energy 221, 386-405

[P2] The hyperparameter space

- Input selection: 11 binary parameters – inclusion of groups of inputs
- A "batch" evaluation during hyperparameter optimization (not a rolling window)
- Bayesian optimization algorithm

[P3] Sources of diversity of the base models

In the paper: random selection of the validation set and data augmentation (4 variants in total)

Other possible methods:

- running multiple independent hyperparameter optimization trials
- running on subsets of data
- augmenting the data using other, similar market (e.g. Belgian data for French model)

Question: does a large ensemble of weak learners outperform a small ensemble of better individual models?

[P3] Speed of the model and optimization time

- The NBEATS-X model: ca. 60% slower to train than the DNN.
- Roughly two days for a hyperparameter optimization run...
- ...but running in minutes when used day-to-day
- The models are fast enough to be used in production

The computational complexity only affects the process of finding a well-performing model

[P3] NBEATSx model extension – basis functions

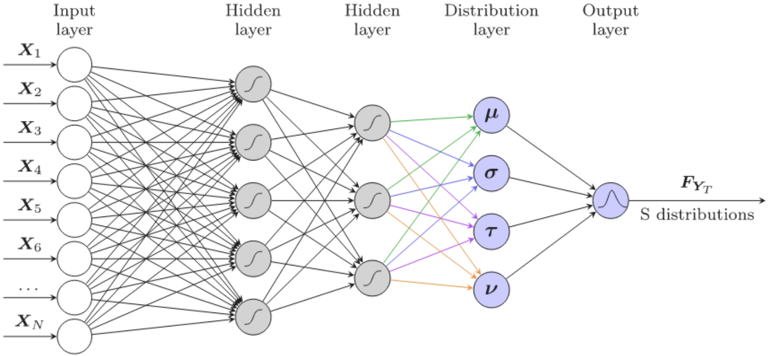
- Using wavelets instead of harmonics would allow to represent a wider variety of functions
- Smoothing can be added to the exogenous variables and the non-interpretable method

[P4] Regularization of bias and neuron response

The main motivation: a separate regularization for the distribution parameter layers

The effect: most hyperparameter optimization trials did not use regularization

The models in the paper did not allow for the bias regularization



[P4] Increasing the complexity – the computational cost

- The proposed method is quick enough for the use in daily operations
- The optimization was not repeated and took a couple days per run (4 runs)
- The profit depends strongly on the cost (trading, battery maintenance) and the scale of operations

	90%	80%	70%	60%	50%	unl_{median}
Naive	7.26					
LEAR-QRA	13.47	12.68	12.28	12.16	12.38	10.40
LEAR-QRM	13.13	12.57	**12.31**	**12.21**	**12.53**	**10.83**
DNN-QRA	13.77	12.26	11.54	11.39	11.37	10.41
DNN-QRM	13.69	11.88	11.57	11.36	11.46	10.56
DDNN-N-pEns	13.99	12.33	11.92	11.78	11.65	9.61
DDNN-N-qEns	13.33	11.93	11.82	11.74	11.64	9.68
DDNN-JSU-pEns	**14.92**	**12.96**	12.15	11.90	11.66	10.37
DDNN-JSU-qEns	14.04	12.45	12.05	12.02	11.79	10.60

[P5] Importance of the exogenous inputs: forecasts

The models typically use the day-ahead forecasts of consumption and RES production:

- this information is avaliable at the time of the auction
- actuals can be used as a proxy (for historical data), but not forecasting
- lagged actuals – infeasible (weather)
- the forecasts for most European markets are publicly available (e.g. on ENTSO-E Transparency platform)

1. Armstrong et al.: Golden rules of forecasting

Due to the nature of EPF, not all rules are applicable

The most relevant rules:

- Use all important knowledge and information by selecting evidence-based methods validated for the situation
- Provide full disclosure for independent audits, replications, extensions
- Use prior knowledge to specify variables, relationships, and effects
- Combine forecasts from diverse evidence-based methods
- Combine forecasts from dissimilar models

1. Armstrong et al.: Golden rules of forecasting

The rules that are not applicable to (short-term) EPF:

- Avoid bias by concealing the purpose of the forecasts
- Most of the judgmental methods section
- Extrapolation methods section – modyfing long-term assumptions about trend and seasonality

2. Model complexity and its performance

The DNN benefits mainly from:

- better description of the non-linear exogenous inputs on the output
- ability to model the data using complex dependencies (e.g. a factor that influences the price up on certain occasions and down otherwise)

3. Fairness of comparing DNN with LASSO

1. LASSO model is single-output, DNNs can be multi-output
    - this property uses the LASSO automated feature selection better – we allow for 24 models tailored for a specific hour
    - training single-output DNNs is possible, but (limited) tests suggest it is better to train a single model (need for a lot of training samples)
2. The DNN model is difficult to fine-tune, LASSO requires very little to work well
    - LASSO offers automatic feature selection, and has only 1 hyper-parameter
    - individual DNN runs are often worse or at par with LASSO – but DNNs can benefit more from averaging
3. DNNs, as a more complex model, *should* outperform LASSO – this is a good verification of the correct implementation