Charles University in Prague
Faculty of Social Science
Institute of Economic Studies

Master's thesis

Comparison of different models for forecasting of Czech electricity market

Bc. Ing. Vladimír Kunc

Supervisor: doc. PhDr. Ladislav Krištoufek, Ph.D.

Study Programme: Ekonomie a Finance

Specialization: FFM&B

July 31, 2017

Declaration

The author hereby declares that he compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain a different or the same degree.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis document in whole or in part.

In Prague on July 31, 2017 Vladimír Kunc ....

Bibliographic note

Kunc, Vladimír. *Comparison of different models for forecasting of Czech electricity market.* 80 p. Master's thesis. Charles University, Fakulty of Social Sciences, Institute of Economic Studies. Supervisor PhDr. Ladislav Krištoufek, Ph.D.

Abstract

There is a demand for decision support tools that can model the electricity markets and allows to forecast the hourly electricity price. Many different approach such as artificial neural network or support vector regression are used in the literature. This thesis provides comparison of several different estimators under one settings using available data from Czech electricity market. The resulting comparison of over 5000 different estimators led to a selection of several best performing models. The role of historical weather data (temperature, dew point and humidity) is also assesed within the comparison and it was found that while the inclusion of weather data might lead to overfitting, it is beneficial under the right circumstances. The best performing approach was the Lasso regression estimated using modified Lars.

JEL Classification C32, C45, C52, C53, Q47

Keywords electricity price forecasting, model comparison, neural networks, support vector regression, kernel ridge regression, lasso, random forest, Diebold–Mariano

Author's e-mail `vlada.kunc@gmail.com`

Supervisor's e-mail `ladislav.kristoufek@fsv.cuni.cz`

Range of thesis: 152,532 characters or 26,267 words.

Abstrakt

Mnoho rozdílných přístupů jako jsou umělé neuronové sítě nebo SVR bývá použito v literatuře. Tato práce poskytuje srovnání několika rozdílných metod v jednotných podmínkách za použití dat z Českého trhu s elektřinou. Výsledné srovnání více jak 5000 modelů vedlo k vybrání několika nejlepších modelů. Tato práce také vyhodnocuje roli historických meteorologických dat (teplota, rosný bod a vlhkost) — bylo zjištěno, že třebaže použití meteorologických může vést k přeučení, za vhodných podmínek může také vést k přesnějším modelům. Nejlepší testovaný přístup představovala Lasso regrese.

Klasifikace JEL C32, C45, C52, C53, Q47

Klíčová slova předpověď cen elektřiny, srovnání modelů, neuronové sítě, SVR, kernel ridge regression, lasso, náhodný les, Diebold–Mariano

E-mail autora `vlada.kunc@gmail.com`

E-mail vedoucího práce `ladislav.kristoufek@fsv.cuni.cz`

Rozsah práce: 152 532 znaků nebo 26 267 slov.

Contents

List of Figures

List of Tables

Master's Thesis Proposal

Institute of Economic Studies
Faculty of Social Sciences
Charles University in Prague

Author:	**Bc. et Bc. Vladimír Kunc**	Supervisor:	doc. PhDr. Ladislav Krištoufek, Ph.D.
E-mail:	85974166@fsv.cuni.cz	E-mail:	ladislav.kristoufek@fsv.cuni.cz
Phone:	774458748	Phone:	
Specialization:	FFM&B	Defense Planned:	June 2017

Proposed Topic:

Comparison of different models for forecasting of Czech electricity market

Motivation:

The electricity markets, due to the nature of the traded good, are one of the most complex markets at the moment. The complexity is caused by several factors that create a combination unique for electricity markets – it is impossible to store electricity efficiently (i.e. economically), the power stability of the grid requires constant balance between production and consumption in real time, and the grid is limited by transmission limits. Furthermore, the production of electricity strongly depends on environmental conditions such as wind speed, temperature, precipitation level or the level of cloudiness which interconnects partially electricity markets with the very complex field of weather prediction and in turn makes sometimes impossible to predict the future production constraints. The electricity markets are also closely related to fuel markets, especially to gas or coal markets, as such fuels are often used in place of electricity (e.g. heating) or for electricity generation (e.g. fossil-fuel power stations).

Likewise, the demand side of the markets is very complex in nature – the demand is influenced by weather conditions as well but it is also influenced by the intensity of business and everyday activities, i.e. there are strong differences in demand during weekdays and weekends, on-peak and off-peak hours, which sometimes results in short-lived, abrupt, and generally unexpected spikes (Wang 2007, Weron 2014). Moreover, there is not one electricity market but rather several interconnected electricity markets in different economical areas. Also, these markets are usually oligopolies where individual entities can and do exert market power (Weidlich 2008, Guerci 2010).

Due to the complexity of the electricity markets, there is a strong demand for decision support tools that can model such markets. There are several different approaches and each with different advantages and disadvantages and it is quite difficult to select the correct tool.

Hypotheses:

The goal of this work is to compare several different model classes and their parametrizations and select the most suitable model for electricity price and demand prediction at the Czech electricity market.
The following hypotheses will be considered:

1. Models that use weather information are more accurate than models that do not
2. Neural network models are more accurate than classical regression models
3. Regression forests are able to perform similarly as other commonly used models

Methodology:

Several different parametrizations of different model classes will be fitted to available data from Czech electricity market. The fitted models will be evaluated on the out-of-sample data for unbiased estimate of performance. The models will be compared across different classes and the several criteria will be used to select several top performing candidates which will be further compared using pairwise Diebold-Mariano test to select the most suitable model for Czech electricity market. Individual approaches will be tested whether the inclusion of weather data allows for more accurate forecasts on the out-of-sample data.

Expected Contribution:

There are many different models and approaches presented in the literature, however the models are usually not compared with each other or when they are, the authors usually compare their model with a basic model from other model class and then generalize that their model is better than models from that class. Thus the goal of this work is to compare many different models under the same settings and to find a model that might be used for forecasting Czech electricity market.

Outline:

1. Electricity market description and motivation – this section describes how electricity markets work and how to they differ from other markets and also why we need a tool for modeling such markets.
2. Literature review and description of common approaches to modeling electricity markets and comparisons of advantages and disadvantages of individual approaches.
3. Model description – this section describes the individual models and how they are estimated.
4. Data – The historic real-world data from the analyzed markets. It will describe how the data were obtained, their source and most importantly, what the data tells us about the market. It will also state which part of the data was used for the estimation of models and which was used for the validation of models.
5. Models evaluation, Results – The models will be evaluated and compared with each other, several best performing candidates will be selected for more detailed analysis
6. Concluding remarks – I will summarize the work and also briefly discuss its implication for future research.

Core Bibliography:

N. Amjady. “Day-Ahead Price Forecasting of Electricity Markets by a New Fuzzy Neural Network”. In: IEEE Transactions on Power Systems 21.2 (May 2006), pp. 887–896. doi: 10.1109/tpwrs.2006.873409. url: http://dx.doi.org/10.1109/TPWRS.
2006.873409.

Bunn, Derek W. and Christopher J. Day (2009). “Computational modelling of price formation in the electricity pool of England and Wales”. In: Journal of Economic Dynamics and Control 33.2, pp. 363–376. doi: 10.1016/j.jedc. 2008.06.009. url: http://dx.doi.org/10.1016/j.jedc.2008.06.009.

Cincotti, Silvano and Giulia Gallo (2013). “The Genoa Artificial Power-Exchange”. In: Communications in Computer and Information Science. Springer Science Business Media, pp. 348–363. doi: 10.1007/978-3-642-36907-0_23. url: http://dx.doi.org/10.1007/978-3-642-36907-0_23.

Möst, Dominik and Massimo Genoese (2009). “Market power in the German wholesale electricity market”. In: JEM 2.2, pp. 47–74. doi: 10.21314/jem. 2009.031. url: http://dx.doi.org/10.21314/JEM.2009.031.

Pinto, Tiago, Tiago M. Sousa, Hugo Morais, Isabel Pra¸ca, and Zita Vale (2016). “Metalearning to support competitive electricity market players’ strategic bidding”. In: Electric Power Systems Research 135, pp. 27–34. doi: 10.1016/j.epsr.2016.03.012. url: http://dx.doi.org/10.1016/j.epsr. 2016.03.012. 1

Tiguercha, A., A.A. Ladjici, M. Boudour, and M. Hazli (2016). “Day ahead a electricity market analysis through a neuroevolution algorithm”. In: 2016 IEEE International Energy Conference (ENERGYCON). Institute of Electrical and Electronics Engineers (IEEE). doi: 10.1109/energycon.2016. 7513907. url: http://dx.doi.org/10.1109/ENERGYCON.2016.7513907.

Ventosa, Mariano, Alvaro Baıllo, Andres Ramos, and Michel Rivier (2005). “Electricity market modeling trends”. In: Energy Policy 33.7, pp. 897–913. doi: 10.1016/j.enpol.2003. 10.013. url: http://dx.doi.org/10.1016/j.enpol.2003.10.013.

Wang, QingQing, ZhaoYang Dong, Xue Li, Junhua Zhao, and Kit Po Wong (2007). “An Effective Approach to Predicting Electricity Market Price Spikes”. In: 2007 IEEE Power Engineering Society General Meeting. Institute of Electrical & Electronics Engineers (IEEE). doi: 10.1109/pes.2007.385852. url: http://dx.doi.org/10.1109/PES.2007.385852.

Weidlich, Anke and Daniel Veit (2008). “A critical survey of agent-based wholesale electricity market models”. In: Energy Economics 30.4, pp. 1728–1759. doi: 10.1016/j.eneco.2008.01.003. url: http://dx.doi.org/10.1016/j.eneco.2008.01.003.

Weron, Rafa l (2014). “Electricity price forecasting: A review of the state-of-theart with a look into the future”. In: International Journal of Forecasting 30.4, pp. 1030–1081. doi: 10.1016/j.ijforecast.2014.08.008. url: http://dx.doi.org/10.1016/j.ijforecast.2014.08.008.

Zhang, Jianhua, Jian Han, Rui Wang, and Guolian Hou. “Day-ahead electricity price forecasting based on rolling time series and least square-support vector machine model”. In: 2011 Chinese Control and Decision Conference (CCDC). Institute of Electrical & Electronics Engineers (IEEE), May 2011. doi: 10.1109/ccdc.2011.5968342. url: http://dx.doi.org/10.1109/CCDC.2011.5968342.

Author **Supervisor**

Chapter 1

Introduction

The energy markets are undergoing significant changes in the last two decades as the energy industry itself has undertaken significant restructuring [7]. Before this restructuring, the electricity markets were usually characterized by monopoly–based organizational structures [7, 138, 169]. The situation has changed in the two last decades and the energy markets resemble free and competitive market now [14, 169]. However, while these markets are already well established, they are still rapidly changing together with the rapid changes in the underlying energy industry such as the increasing share of renewables or the planned introduction of smart grids.

However, while the commodity markets has been around for decades, the knowledge from these markets cannot be simply transferred to the energy markets as the electrical energy differs significantly from most other commodities and thus the energy markets has their own peculiarities. First of all, the electrical energy cannot be efficiently stored in the necessary amounts [197]. Furthermore, in order to ensure the stability of the power system, constant balance between supply and demand is required [197]. Furthermore, the production of electrical energy is not as easily controllable as the most of types of power-plants cannot be regulated within short time interval, thus it is not possible to just shutdown the necessary amount of power-plants when the supply is higher than the demand. This trait of the energy industry leads to, among other things, the emergence of negative electricity prices on the market — a phenomenon that is not common in the other commodity markets. Moreover, the electrical markets are very complex not only because of the impossibility of economical storage of electricity but also because the cost of electricity generation varies with the weather — there is a not insignificant amount of wind and solar power-plants in the industry.

However, the energy markets are not complex only because of the supply side of the energy industry but also because the demand of electricity depends on many things. First of all, the demand depends on the structure of business intensity — there is a significant difference in electricity demand of some consumers in on–peak hours and off–peak hours and also during the weekdays and the weekends). Furthermore, even the demand strongly depends on the weather as, for example, people might use electricity for heating when the weather is cold or use it for air–conditioning when the weather it hot. Thus both electricity demand and supply are changing with the seasons. This dependency result in complex market behavior and the electricity price exhibits both intraday periodic cycles and more long term cyclical changes such as weekly changes or even seasonal changes, which together with the changing energy industry, makes the electricity markets quite complex and hard to predict their behavior in both short-term and long-term.

Moreover, the market behavior does not only exhibit seasonalities at various levels (daily, weekly, or annually) but it also exhibits sudden and hard to predict spikes [195, 197]. While the spikes themselves are actively researched, there is still no general consensus about the cause behind them and they are still very hard to predict [219]. The understanding of the causes of these spikes is a very important task for risk management of market participants.

This work aims to compare many different models that are commonly used for electricity price prediction on the Czech electricity market operated by OTE, a.s.[1] as no extensive model comparison for short-term electricity price forecasting is present in the literature (with the exception of [24] which compares several different models — more in Ch. 2). While electricity price prediction is actively researched topic in the literature, almost none of the works focuses on the Czech electricity market which is quite different environment compared to other, more frequently researched markets — for in depth research about the Czech electricity market and its regulation, see [100]. The goal of this work is to compare approaches that can be directly utilized for trading thus this work focuses on hourly price prediction two days ahead — more specifically, all used methods forecast a set of 24 hourly prices for day $t + 2$ at time t. This allows to make bids in the market as the market closes the previous day. The $t + 2$ forecast is a bit more challenging than the $t + 1$ (which is not usable for trading on the Czech electricity market) as it introduces more uncertainty and thus the qualitative results from this work cannot be directly compared to works that utilize a $t + 1$ forecasts.

This work focuses on artificial neural networks models, regression models and other machine learning models — while it is nearly impossible to cover all possible models, this thesis tries to cover various parametrizations of the most common models — this works compares over 5000 estimators and their parametrizations. Despite the number of compared estimators , agent–based models (ABMs) are not used in this works — however, most of the predictive ABMs utilize internally an ANN or SVM for the price forecasting (e.g. [144, 145, 169]) thus the results of this thesis might be used in the ABMs.

The thesis has the following structure. A brief review of related work is provided in chapter 2. The used methods are presented in chapter 3. The description of PSO and NM methods used for optimization of estimators' parameters is provided in section 3.1, the individual estimators that were used in this thesis are described in section 3.2, and methods used for evaluation of the forecasts are presented in section 3.3. Used data are described in chapter 4 — including the individual seasonalities present in the data. And finally, the used estimators are compared in chapter 5. Concluding remarks and the possible extension of this work are in chapter 6.

[1] `http://www.ote-cr.cz`

Chapter 2

Related works

There has been a significant increase in the literature about electricity price forecasting in the last decade. The research of short-term electricity price prediction has begun in 90s and then the number of published works rocketed in 2000s— there were 206 Scopus indexed journal articles and 274 conference papers in 2013 [197]. There are two main classes of commonly used models — *statistical models* (e.g. *OLS regression*, *ARIMA* models) and *artificial neural networks models* (ANNs). However, the distinction between these two classes is blurry as there are models that can fit both classes, e.g. linear regression using OLS is equivalent to a ANN with a single neuron with linear activation function — the models are the same only the common optimization approaches differs (OLS has a closed-form solution while ANN usually uses some kind of iterative numerical optimization such as *gradient descent*). Furthermore *support vector regression* (SVR) is also sometimes considered to be a kind of neural network albeit with completely different structure and optimization compared to classical *feed forward* neural networks [188].

It is greatly out of the scope to discuss all the related literature, thus only the most relevant works are briefly described. A sample of related works for last two decades is provided in table 2.1. While the sample is not random, it can be observed that two most popular approaches are ANNs and ARIMAs models and their modifications. A third quite popular approach is based on SVR (by some considered to be an ANN). Other approaches such as *decision trees* (DTs) [47, 148], *random forests* (RFs) [57, 122], *local informative vector machines* (IVMs) [43], *relevance vector machines* [4] are less frequent but still usable and able to provide quite accurate forecasts. The list of used methods is incomplete and it is just to show the variety of approaches used for modelling electricity prices as almost every work about electricity price prediction introduce a novel method or a novel modification of an existing method.

The most popular ANNs for electricity price prediction are feed forward ANNs. A feed forward ANN with a single hidden layer with 8–12 neurons was used in [155]. Two layered feed forward ANN was used in [171] for day ahead price prediction in the New England electricity market. A feed forward ANN was also used in [208] where its performance was compared to other popular estimators, namely the SVR, RFs and K-nearest neighbor (KNN) based models. One of the rare works about the Czech electricity market is [172] where the authors compare an ANN model with

ARIMA and hybrid ARIMA-ANN models for predicting the spot prices. While the hybrid model performed better than the pure ARIMA models it was still outperformed by pure ANN model in their settings [172]. Similar results were obtained in another work focused on the Czech electricity market [24] which compared several different model for prediction of both hourly and daily prices. An ANN model (*multi-layer perceptron*) performed the best for prediction of hourly prices (RMSE 6.15 EUR/MWh) while an ARIMA model gave slightly worse prediction (RMSE 6.25 EUR/MWh). However, even worse prediction (RMSE 7.32 EUR/MWh was made by another popular ANN model — *Elman networks*. The author also reports different results when the used metrics was *mean absolute percentage error* (MAPE) as the ARIMAX was worse than the *Elman network*. However, this was probably caused by the unsuitability of MAPE as this metrics is biased towards lower forecasts [184] due to its asymmetry [92, 163]. Another work focused on the Czech electricity market is [102] which, however, dealt mainly with the prediction of volatility of electricity prices even though one of the estimators was producing day ahead price predictions during the volatility forecasting process [102].

SVR and SVMs are quite popular as they are able to deal well with outliers and are able to provide quite robust forecasts [206] (even though even more robust versions have been proposed in [80, 84, 186, 187]) and as SVRs can be used instead of classical linear regression without any major changes in workflow. Furthermore, the kernelized versions of SVR and SVMs allow implicit mapping to different space and thus can be used even for non-linear problems. The parameters of SVR for electricity price prediction were optimized using PSO in [165, 166] to tune the regularization of the regression.

DTs and RFs represent completely different approach than LR and SVR — they fit a non-continuous function to the data by creating decision tree with thresholds. While this approach might not be intuitive for regression problems it can deal with missing data [114] and the decision trees are easily interpretable [114]. A DTs and RFs can function as a black–box algorithm as they can be applied directly to the data without any pre-processing or assumptions. DTs were used for electricity price prediction in [47] where each prediction is done by 24 models, each predicting one particular hour in the day-ahead prediction. The first step is classification of individual datapoints as either regular data point or a spike. Second step is a feature selection process. Then finally the models for price prediction are fitted — each of the 24 models consists of 3 DTs. First DT predicts whether a spike will occur, the second model predicts the price if the spike occurs and the third if the spike does not occur. The advantage of such approach is that it also facilitates spike prediction. DTs might be used also as part of feature selection process before yielding the prediction itself to another estimator as in [148] where the author studies two different algorithms for fitting DTs (ID3 and C4.5) for feature selection for price prediction.

Table 2.1: Summary of related works.

Sample or related works

year	work	method	market	notes
1998	[194]	ANN	England and Wales	
1999	[13]	expert-based model	JPM	*MAPS* software
1999	[175]	ANN	Victoria	
2000	[136]	ANN	—	
2000	[35]	expert-based structural model	California	
2000	[210]	ANN	UK power pool	using *wavelet transform* and RBF NN
2000	[50]	ANN	California	
2001	[135]	AR with classification	New England	estimating also PDF
2001	[90]	covariance TS model	California	
2002	[138]	Dynamic regression	Spain, California	

Continuation of Table 2.1				
year	**work**	**method**	**market**	**notes**
2002	[76]	ANN	JPM	
2002	[137]	ANN	California	
2002	[91]	OLS	UK power pool	using *wavelet transform*
2002	[117]	KNN, Dynamic regression	Spain	
2003	[30]	ARIMA	Spain, California	
2003	[64]	ANN	New England	using *RBF* networks
2003	[217]	ANN	New England	using *Cascaded NN*
2003	[209]	ANN	New England	using *Elman network*
2004	[205]	ARMA	JPM	using *wavelet transform*
2004	[119]	ANN	JPM, New England	
2004	[221]	ARIMA	California	
2004	[207]	ANN	California	
2004	[79]	ANN, LR	China	both short–term and long–term forecast
2004	[65]	ANN ensemble	New England	
2004	[32]	ARMA	Leipzig Power Exchange	comparison of differnent ARMA models
2004	[75]	ANN	JPM	using *recurrent NN* with *fuzzy* rules
2004	[155]	ANN	Ontario	
2004	[203]	ANN	China	using *wavelet NN*
2005	[29]	ARIMA	Spain	using wavelet transform
2005	[123]	Hidden Markov Models	Spain	
2005	[198]	ARMAX	California	
2005	[118]	ANN	ANEM	with Bayesian classification for spike prediction
2005	[216]	ANN	New England	using *Kalman filter*
2005	[51]	ARMA,GARCH	California, Spain	
2006	[220]	ARIMA	California	
2006	[54]	ANN	Leipzig Power Exchange [2]	
2006	[142]	ANN,AR	Leipzig Power Exchange	predicting daily prices
2006	[5]	ANN	Spain	using *fuzzy NN*
2006	[56]	ANN	California	
2006	[202]	Bayesian Expert with SVM	China	
2006	[109]	ANN	JPM	trained using *artificial fish swarm*
2006	[111]	ANN	California	*recurrent NN*
2007	[116]	ANN, KNN	Spain	
2008	[124]	ARMA	NordPool, Ontario	
2008	[78]	ANN, LR	China	
2008	[139]	SARIMA, GARCH	Spain	
2010	[201]	ANN,ARMAX,GARCH	PJM	hybrid model *adaptive wavelet neural network*
2010	[108]	Grey model	NordPool	PSO tuned
2010	[6]	ANN ensemble	Italy, New England, Ontario	
2010	[176]	ARIMA, GARCH	Spain, PJM	using *wavelet transform*
2010	[180]	ANN	PJM	using *extreme learning machines*
2010	[178]	ANN ensemble	PJM	using *extreme learning machines bagging*
2010	[179]	ANN ensemble	PJM	using *extreme learning machines ensemble*
2010	[58]	GARCHX	UK power pool	daily prices
2010	[115]	EWM	NordPool, France, The Netherlands	
2010	[147]	ANFIS	Spain	using wavelet fuzzy NN
2011	[177]	ANN	PJM	using *extreme learning machines*, *RBF NN*
2011	[171]	ANN	New England	
2011	[9]	ANN	PJM	
2011	[170]	ANN	New South Wales	both price and load forecasting
2011	[127]	LR	Iran	trained using GA
2011	[215]	ANN	—	
2011	[214]	ARMAX,GARCH, LS–SVM	California	hour ahead
2011	[165]	PSO tuned SVR	Ontario	
2012	[208]	SVR,ANN,RF,KNN, LR	—	daily prices, comparison of various approaches
2012	[4]	RVM	New England	
2012	[52]	Dynamic Factor Model	Spain	
2012	[149]	SVR	ANEM, PJM, Spain	
2012	[185]	SVR, ANN	NordPool, Ontario	both wavelet and RBF NN
2012	[132]	ANN, NARX	New South Wales, New England	using *data association mining*
2013	[121]	ANN	Ontario	with feature selection
2013	[83]	ARIMAX ensemble	NordPool	
2013	[43]	IVM	Spain	comparison with other methods
2013	[204]	SARIMA	NordPool	

Continuation of Table 2.1

year	**work**	**method**	**market**	**notes**
2013	[200]	RDFA	ANEM	
2013	[87]	Multi-Kernel learning	MISO	
2014	[148]	C4.5, ID3	New South Wales, Spain	
2014	[166]	PSO tuned SVR	Ontario	interval estimation
2014	[140]	evolutionary based ANN	Portugal	3 hours ahead
2014	[86]	Multi-Kernel Learning	MISO	
2014	[131]	ARMA, SVM	Italy	
2014	[17]	ANN	NordPool	focus on wind power bidding strategy
2014	[62]	SETAR	Italy	
2014	[193]	ANN	ANEM	hybrid ELM
2014	[12]	ANN, ANFIS, ARMA	Spain	using *Kalman Fusion*
2015	[1]	ANN	PJM	Combinatorial NN
2015	[168]	ANN	Ontario	multiobjective, interval estimation
2015	[167]	PSO tuned SVR	Ontario,PJM	interval estimation, comparison
2015	[164]	SVR	Germany	using natural gas prices
2015	[172]	ARIMA,NN	Czechia	hybrid ARIMA-NN model
2015	[47]	DT	Belgium, NordPool	spike prediction
2015	[63]	STAR	Italy	
2015	[95]	ANN with PSO	Greece, U.S. cities	wavelet NN
2015	[96]	LR	Nordpool	
2015	[57]	RF, ANN	Spain	
2015	[122]	RF, LS–SVM	Victoria	
2015	[18]	ANN	ANEM, Spain	*fuzzy NN*
2016	[141]	ANN	SUD Italy	
2016	[99]	ANN	UK power pool, Serbia	both short–term and long–term
2016	[8]	ANN, LR	India	daily prices
2016	[153]	LSSVM-GA	Ontario	
2016	[34]	ANN, SVR	Italy	
2016	[130]	SVR	New England	daily prices
2016	[129]	SVR	New England	daily prices, mid–term forecast
2016	[128]	SVR	New England	daily prices, feature selection
2016	[189]	Markov Switch Model	NordPool	using GARCH,MRJD
2016	[106]	ANN, SVR, RF	New South Wales	
2017	[192]	ANN with GA	Victoria	ELM with NSGA-II, interval estimation
2017	[53]	Mixed models, SARIMA	Spain	also general introduction
2017	[151]	ANN	ANEM, Ontario	wavelet neural network and generalized ELM

End of Table 2.1

Chapter 3

Methods

Reviews of different methods for electricity price forecasting are available in [2].

3.1 Optimization of hyperparameters

While individual regression algorithms have their own methods for minimizing the error of regression, most of those algorithms are parametrized by hyperparameters. Hyperparameters are parameters that influence the function of a regression algorithm but are not optimized in the algorithm itself. Commonly used hyperparameters are regularizing constants or, in case of ensembles, number of estimators within the ensemble.

This work uses several approaches for selection of the suitable hyperparameters for individual algorithms. The hyperparameters were optimized by training different parametrization on the *training set* and then evaluated using the *validation set*. Three different methods were used for directing the search in the space of possible hyperparameters: grid search (GS) [16], particle–swarm optimization (PSO) [19, 89, 146] and the Nelder–Mead (NM) method [105, 126, 134].

3.1.1 Grid search

Grid search is an undirected search of the optimization space that explores points spaced evenly in all dimensions. This method is conceptually very simple and very easy to parallelize as coordinates of each points are calculated at the beginning of the search and then only independently evaluated. The GS is not suitable for high-dimensional optimization as it suffers from the curse of dimensionality [16] — the number of points to evaluate grows exponentially with the dimension. Furthermore, the GS does not work well for functions with *low effective dimensionality* (i.e. some dimensions are much more important than the others) in high dimensional search space — for details viz [16]. Due to these problems, the GS method is used within this work only for quite small problems.

3.1.2 Particle–swarm optimization

Particle–swarm optimization (PSO) is a stochastic search population based method that iteratively searches the optimization space [89]. PSO has been successfully used in various optimization problems (for reviews of application viz [19, 146]) as it is both simple and robust class of methods that are performing well even for complicated functions.

3.1.2.1 Usage

This work uses PSO for finding optimal values of hyperparameters for a representative estimator of a class of estimators. The found optimal values of hyperparameters of the representative estimator are then used as an initial point for local search for other estimators from the class.

3.1.2.2 Algorithm

Tere are several different PSO algorithms that have been proposed over last two decades (viz [19]), however the used PSO algorithm is based on [28] and is described below.

The method uses a population of m particles that are iteratively updated and explore n dimensional space that is constrained by initial range $c_{j_{\min}}$ and $c_{j_{\max}}$ for $j \in \{1, \ldots, n\}$. Each particle i at iteration t has defined [19, 28]:

- position $\boldsymbol{x}_i^t$

  Position describes the coordinates in the search space and is directly responsible for the quality of a particle.

- velocity $\boldsymbol{v}_i^t$

  Velocity describes the direction and length of movement of a particle.

- personal best $\boldsymbol{p}_i^t$

  Personal best describes the best position that a particle visited in since beginning of the algorithm.

The algorithm then iteratively updates the particles

1. the positions $\boldsymbol{x}_i^0$ and velocities $\boldsymbol{v}_i^0$ are randomly initialized within the constrained initial range, $p_i^0 = \mathrm{f}(\boldsymbol{x}_i^0)$ for each particle $i \in \{1, \ldots, m\}$

2. find the historical minimum $\boldsymbol{p}_g^t$ over all particles at time t

$$\boldsymbol{p}_g^t = \operatorname*{arg\,min}_{i \in \{1,\ldots,m\}} \mathrm{f}(\boldsymbol{p}_{\boldsymbol{i}}^{\boldsymbol{t}}) \tag{3.1}$$

3. then for each particle

(a)

$$\boldsymbol{v}_i^{t+1} := \boldsymbol{v}_i^t + \boldsymbol{\phi_1} \cdot (\boldsymbol{p}_i^t - \boldsymbol{x}_i^t) + \boldsymbol{\phi_2} \cdot (\boldsymbol{p}_g^t - \boldsymbol{x_i^t}) \tag{3.2}$$

(b)

$$\boldsymbol{x}_i^{t+1} = \boldsymbol{x}_i^t + \boldsymbol{v}_i^{t+1} \tag{3.3}$$

(c)

$$\boldsymbol{p}_i^{t+1} := \min(\boldsymbol{p}_i^t, \mathrm{f}\left(\boldsymbol{x}_i^{t+1}\right) \tag{3.4}$$

where $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ are vectors of n uniformly sampled values in $[0, \phi_{1_{\max}}]$ and $[0, \phi_{2_{\max}}]$.

4. repeat 2 and 3 until termination criterion

3.1.3 Nelder–Mead

Nelder–Mead simplex method is a numerical optimization method for non-linear problems introduced in [134]. The method search a n dimensional space using a $n+1$ vertices arranged in a simplex. The method computes the objective function at each vertex of the simplex and then extrapolates the objective function to find a better point in the search space that subsequently replaces the vertex with the worst objective value in the simplex and then repeat the extrapolation.

While the Nelder–Mead method works well for many functions, it might converge to a non-stationary point of the objective function [126]. The Nelder–Mead method might still be useful despite the possible convergence to a non-stationary point because it requires only a relatively few evaluations of the objective functions [134].

3.1.3.1 Usage

Since this Nelder–Mead method might have convergence problems, this method is used in this work only as a possible refinement when the starting point should already be close to an optima — the PSO algorithm is used to find optimal parameters of a representative of a class of estimators and then the Nelder-Mead is used for possible update of the optimal parameters of the representative estimator for each estimator from the class. Since the estimators within the class are similar, their optimal values of hyperparameters are mostly close as well. The Nelder–Mead method is thus used only as a heuristic local search and therefore the convergence problem are not an issue.

3.1.3.2 Algorithm

The Nelder–Mead method is an iterative approach that updates a simplex of $n+1$ points in n-dimensional search space in each iteration [105]. The minimization version from [105] is described below:

1. initialize the simplex of $n+1$ points: $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n, \boldsymbol{x}_{n+1}\}$

2. order and reindex the points by respective values of objective function

$$\mathrm{f}(\boldsymbol{x}_1) \leq \mathrm{f}(\boldsymbol{x}_2) \leq \ldots \leq \mathrm{f}(\boldsymbol{x}_n) \leq \mathrm{f}(\boldsymbol{x}_{n+1}) \tag{3.5}$$

3. calculate centroid $\overline{\boldsymbol{x}}$ of points $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, $\boldsymbol{x}_3$,..., $\boldsymbol{x}_{n-1}$, $\boldsymbol{x}_n$

4. Compute the *reflection point* $\boldsymbol{x}_r$ and evaluate $\mathrm{f}(\boldsymbol{x}_r)$

$$\boldsymbol{x}_r := \overline{\boldsymbol{x}} + \rho(\overline{\boldsymbol{x}} - \boldsymbol{x}_{n+1}) \tag{3.6}$$

5. If $\mathrm{f}(\boldsymbol{x}_1) \leq \mathrm{f}(\boldsymbol{x}_r) < \mathrm{f}(\boldsymbol{x}_n)$, continue to 2 with $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n, \boldsymbol{x}_r\}$

6. If $\mathrm{f}(\boldsymbol{x}_r) < \mathrm{f}(\boldsymbol{x}_1)$, calculate the *expansion point* $\boldsymbol{x}_e$ and evaluate $\mathrm{f}(\boldsymbol{x}_e)$

$$\boldsymbol{x}_e := \overline{\boldsymbol{x}} + \chi(\boldsymbol{x}_r - \overline{\boldsymbol{x}}) \tag{3.7}$$

   If $\mathrm{f}(\boldsymbol{x}_e) < \mathrm{f}(\boldsymbol{x}_r)$, continue to 2 with $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n, \boldsymbol{x}_e\}$, else continue to 2 with $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n, \boldsymbol{x}_r\}$

7. If $\mathrm{f}(\boldsymbol{x}_n) \leq \mathrm{f}(\boldsymbol{x}_r) < \mathrm{f}(\boldsymbol{x}_{n+1})$, calculate *outside contraction* point $\boldsymbol{x}_{oc}$ and evaluate $\mathrm{f}(\boldsymbol{x}_{oc})$

$$\boldsymbol{x}_{oc} := \overline{\boldsymbol{x}} + \gamma(\boldsymbol{x}_r - \overline{\boldsymbol{x}}) \tag{3.8}$$

   If $\mathrm{f}(\boldsymbol{x}_{oc}) \leq \mathrm{f}(\boldsymbol{x}_{n+1})$, continue to 2 with $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n, \boldsymbol{x}_{oc}\}$

8. If $\mathrm{f}(\boldsymbol{x}_r) \geq \mathrm{f}(\boldsymbol{x}_{n+1})$, calculate *inside contraction* point $\boldsymbol{x}_{ic}$ and evaluate $\mathrm{f}(\boldsymbol{x}_{ic})$

$$\boldsymbol{x}_{ic} := \bar{\boldsymbol{x}} - \gamma(\boldsymbol{x}_r - \bar{\boldsymbol{x}}) \tag{3.9}$$

   If $\mathrm{f}(\boldsymbol{x}_{ic}) \leq \mathrm{f}(\boldsymbol{x}_{n+1})$, continue to 2 with $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_d, \boldsymbol{x}_{ic}\}$

9. For each point i, $i \in \{2, 3, \ldots, n, n+1\}$ calculate $\boldsymbol{x}'_i$ and evaluate $\mathrm{f}(\boldsymbol{x}'_i)$

$$\boldsymbol{x}'_i := \boldsymbol{x}_1 + \sigma(\boldsymbol{x}_i - \boldsymbol{x}_1) \tag{3.10}$$

   Continue to 2 with $\{\boldsymbol{x}_1, \boldsymbol{x}'_2, \ldots, \boldsymbol{x}'_d, \boldsymbol{x}'_{n+1}\}$.

where ρ, χ, γ, and $\sigma \in \mathbb{R}$ are coefficients of individual operations [105]. The standard values are $\rho = 1$, $\chi = 2$, $\gamma = \frac{1}{2}$, and $\sigma = \frac{1}{2}$ [105].

3.2 Estimators

Various estimators are compared in this work, while this comparison cannot include all different estimators and their parametrizations, nevertheless this thesis attempts to provide comparison of the major and most intuitive classes of estimators. This section contains short overview of each class of estimators that are compared in this work.

3.2.1 Ordinary Least Squares Regression

Ordinary Least Squares (OLS) is probably the most common linear estimator as it is very simple and it works well in many cases. History and properties of the OLS method are out of scope of this work and can be found in any textbook of Econometrics — e.g. [199, p. 72]. Even though the underlying model of hourly electricity prices is not linear and it evolves dynamically in time, the OLS might still be able to provide a good prediction of the prices. The quality of a prediction is evaluated using out-of-sample data.

The OLS tries to fit linear model:

$$y = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta} + \epsilon \tag{3.11}$$

where y is the *dependent* variable and $\boldsymbol{x} \in \mathbb{R}^n$ is vector of n independent variables and $\boldsymbol{\beta} \in \mathbb{R}^n$ is vector of n weights.

The optimization formulation of the problem tries to find such $\widehat{\boldsymbol{\beta}}$ that minimize the squared error:

$$\widehat{\beta} = \arg\min_{\beta} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2 \tag{3.12}$$

where $\|\boldsymbol{x}\|_2 = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}$ for $\boldsymbol{x} \in \mathbb{R}^n$ is the Euclidean (L_2) norm. The OLS is very simple and does not impose other constraints on the estimates $\widehat{\boldsymbol{\beta}}$, which makes OLS sensitive to outliers and overfitting in some settings.

3.2.2 Support Vector Regression

Support Vector Regression (SVR) is a regression method based on the same idea as the famous Support Vector Machines (SVMs) [31] for classification tasks. It was first proposed in [41] and similarly as SVM, the SVR model depends only on a subset of data. This work has used several different formulation of SVRs.

3.2.2.1 ϵ-SVR with linear kernel

The simplest ϵ-SVR is with linear kernel. This work has used two formulations of linear ϵ-SVRs — with L_1 loss (ϵ*-insensitive* and with L_2 loss (*squared* ϵ*-insensitive* loss).

The optimization formulation is:

$$\min_{\boldsymbol{w}} \frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w} + \begin{cases} C\sum_{i=1}^{n}\left(\max\left(0, \left|y_i - \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i\right| - \epsilon\right)\right) & \text{if using } L_1 \text{ loss} \\ C\sum_{i=1}^{n}\left(\max\left(0, \left|y_i - \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i\right| - \epsilon\right)\right)^2 & \text{if using } L_2 \text{ loss} \end{cases} \tag{3.13}$$

where $\boldsymbol{w} \in \mathbb{R}^n$ is vector of n weights, C is the regularization constant and $\epsilon \in \mathbb{R}_{\geq 0}$

is determining the sensitiveness of the loss [46].

The dual form of the optimization is:

$$\begin{aligned}
&\min_{\boldsymbol{a}^+,\boldsymbol{a}^-} \frac{1}{2}\begin{bmatrix}\boldsymbol{a}^+ & \boldsymbol{a}^-\end{bmatrix}\begin{bmatrix}\overline{\boldsymbol{Q}} & -\boldsymbol{Q}\\ -\boldsymbol{Q} & \overline{\boldsymbol{Q}}\end{bmatrix}\begin{bmatrix}\boldsymbol{a}^+\\ \boldsymbol{a}^-\end{bmatrix} - \boldsymbol{y}^{\mathrm{T}}\left(\boldsymbol{a}^+ - \boldsymbol{a}^-\right) + \epsilon\boldsymbol{1}^{\mathrm{T}}\left(\boldsymbol{a}^+ + \boldsymbol{a}^-\right)\\
&\text{w.r.t.}\\
&\qquad 0 \leq \alpha_i^+, \alpha_i^- \leq U, i = 1, 2, \ldots, m
\end{aligned} \tag{3.14}$$

where $\boldsymbol{1}$ is the vector of all ones, $\overline{\boldsymbol{Q}} = \boldsymbol{Q} + \boldsymbol{D}$, $\boldsymbol{Q} \in \mathbb{R}^{m \times m}$ and $Q_{ij} := \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j$, and $\boldsymbol{D}$ is a diagonal matrix defined as:

$$D_{ii} = \begin{cases} 0 \text{ if using } L_1 \text{ loss} \\ \frac{1}{2C} \text{ if using } L_2 \text{ loss} \end{cases} \tag{3.15}$$

and U [46]:

$$U = \begin{cases} C \text{ if using } L_1 \text{ loss} \\ \infty \text{ if using } L_2 \text{ loss} \end{cases} \tag{3.16}$$

The used implementation [46] actually optimize slightly different problem that is equivalent, for details see [46].

3.2.2.2 ϵ-SVR with a general kernel

Other kernels allows the ϵ-SVR implicitly map the problem into different space where the SVR might find better solution similarly as in regular SVM with a kernel.

The optimization problem of an ϵ-SVR with a general kernel is [25]:

$$\begin{aligned}
&\min_{\boldsymbol{w},b,\boldsymbol{\xi},\boldsymbol{\xi}^*} \frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w} + C\sum_{i=1}^{m}\xi_i + \sum_{i=1}^{m}\xi_i^*\\
&\text{w.r.t.}\\
&\qquad \boldsymbol{w}^{\mathrm{T}}\phi\left(\boldsymbol{x}_i\right) + b - y_i \leq \epsilon + \xi_i,\\
&\qquad y_i - \boldsymbol{w}^{\mathrm{T}}\phi\left(\boldsymbol{x}_i\right) - b \leq \epsilon + \xi_i^*,\\
&\qquad \xi_i, \xi_i^* \geq 0, i = 1, 2, \ldots, m
\end{aligned} \tag{3.17}$$

and the dual problem is [25]:

$$\begin{aligned}
&\min_{\boldsymbol{\alpha},\boldsymbol{\alpha}^*} \frac{1}{2} \left(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*\right)^{\mathrm{T}} \boldsymbol{Q} \left(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*\right) + \epsilon \sum_{i=1}^{m} (\alpha_i + \alpha_i^*) + \sum_{i=1}^{m} y_i(\alpha_i - \alpha_i^*) \\
&\text{w.r.t.} \\
&\qquad \mathbf{1}^{\mathrm{T}} \left(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*\right) = 0, \\
&\qquad 0 \leq \alpha_i, \alpha_i^* \leq C, i = 1, 2, \ldots, m
\end{aligned} \tag{3.18}$$

where $Q_{ij} = K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \phi(\boldsymbol{x_i})^{\mathrm{T}}\phi(\boldsymbol{x}_{)}$, i.e. $K(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is determined by the used kernel [25].

The approximate function is [25]:

$$\mathrm{f}(\boldsymbol{x}) = \sum_{i=1}^{m} \left(-\boldsymbol{\alpha} + \boldsymbol{\alpha}^*\right) K(\boldsymbol{x}_i, \boldsymbol{x}) + b. \tag{3.19}$$

3.2.2.3 Used kernels

This works uses several different kernels that are described below, however, the list is limited only to used kernels, for deeper introduction into kernel see [190].

3.2.2.3.1 Linear kernel The linear kernel is the most simple one and is defined as [143, 190]:

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{x}_j \tag{3.20}$$

3.2.2.3.2 Polynomial kernel The polynomial kernel is defined as [143, 190]:

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \gamma \left(\boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{x}_j + c_0\right)^d, \tag{3.21}$$

where $\gamma \in \mathbb{R}_{>0}$ is scaling parameter, c_0 is usually set either to 0 (*homogeneous polynomial kernel*) or 1 (*inhomogeneous polynomial kernel* — other values are also possible). The *homogeneous* kernel with $c_0 = 0$ and $\gamma = \frac{1}{n}$ is used for the purposes of this work.

3.2.2.3.3 Sigmoid kernel This kernel is also known as *hyperbolic tangent kernel* [143, 190]:

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \tanh\left(\gamma \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{x}_j + c_0\right) \tag{3.22}$$

where c_0 and γ has the same meaning as in the *polynomial kernel*. This work uses the kernel with $c_0 = 0$ and $\gamma = \frac{1}{n}$.

3.2.2.3.4 Radial basis function kernel The radial basis function kernel (RBF) [143, 190] is one of the most commonly used kernels with SVMs:

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\gamma \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2\right). \tag{3.23}$$

Sometimes the kernel is defined with σ instead of γ and the usual meaning is $\gamma = \frac{1}{\sigma^2}$ [143] or $\gamma = \frac{1}{2\sigma^2}$ [190].

3.2.3 Decision Trees

Decision Trees (DTs) might be used for both regression and classification — when used for regression, they actually discretize the output into varying levels. The tree consists of binary decision nodes which contain condition and based on the condition, the estimation continues with either left or right branch [114]. The leaves usually contain an output value that is returned.

Decision Trees are very useful as they are very easily and intuitively interpretable [114]. The disadvantage of DTs is that they are prone to overfitting — esp. in case of high-dimensional input and few samples. The DTs are also popular for ensemble methods such as *Random Forests* (RFs) [22] or AdaBoost [48], viz [3, 11, 38, 88, 110] for details of DTs in ensembles.

DTs are a general class of approaches and there is quite a number of algorithms for fitting DTs to data — e.g. AID, THAID, MAID, ELISEE, ID3, CART and C4.5, viz [98, 113, 114, 133, 150, 154, 174] for reviews and details. This work uses the CART algorithm that allows both classification and regression. A DT has many different parameters that strongly influence its shape and performance, e.g. *max features per split* (the maximum number of features considered in each split), *maximal depth* (used often for reducing overfitting), *minimum samples per split* (the minimum samples needed for a split), *minimum samles in a leaf*, *maximum number of leaves*, or *minimum impurity of a split* (for stopping the growth of a tree early), for details about the CART algorithm see [114, 160].

3.2.4 Ridge Regression

Ridge Regression (RR) is very similar to OLS but it adds L_2 regularization on estimated coefficients. The first use of this method for regression was presented in [73]. The optimization formulation of RR is defined as [143]:

$$\min_{\boldsymbol{w}} \|\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}\|_2^2 + \alpha \|\boldsymbol{w}\|_2^2 \tag{3.24}$$

where $\alpha \in \mathbb{R}_{\geq 0}$ is the regularization parameter controlling the amount of the shrinkage.

3.2.5 Lasso Regression

Lasso (*least absolute shrinkage and selection operator*) regression is another extension with OLS that regularize estimated coefficients. The Lasso tries to estimate sparse coefficients (and thus it can be also seen as a tool for feature selection) [181] by regularizing the coefficients with the L_1 norm:

$$\begin{aligned} &\min_{\boldsymbol{w}} \frac{1}{m} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}\|_2^2 \\ &\text{w.r.t.} \\ &\|\boldsymbol{w}\|_1 \leq t \end{aligned} \tag{3.25}$$

where t is the regularizing parameter.

It can be also rewritten to an equivalent problem [143]:

$$\min_{\boldsymbol{w}} \frac{1}{2m} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}\|_2^2 + \alpha \|\boldsymbol{w}\|_1 \tag{3.26}$$

where α is a different regularizing constant and $\|\boldsymbol{w}\|_1 := \sum_{i=1}^{n} |w_i|$ is the L_1 norm.

3.2.6 Least Angle Regression

Least Angle Regression (LARS) is a regression method similar to Lasso that both fits a linear model to data and performs an implicit feature selection similar to *forward selection* [182].

The algorithm works as follows [182]:

1. Start with $\beta_i := 0, i = 1, 2, \ldots, n$
2. find predictor x_{j_1} most correlated with the response
3. take largest possible step (i.e. increase in β_j) until some other predictor x_{j_2} has as much correlation with the current residual
4. form set of predictors S
5. take largest possible step equiangular between predictors from S (e.g. in second iteration S $= \{x_{j_1}, x_{j_2}\}$) until some other predictor x_{j_i} has as much correlation with the current residual
6. add x_{j_i} to S
7. continue with step 5 until all predictors are in the model

for more details about the algorithm and how to compute the equiangular direction effectively, viz [182].

3.2.6.1 Lasso–LARS

The slightly modified LARS allows for computing exact solution for Lasso regression [182] — this method is denoted as Lasso–LARS and was used to compare to regular Lasso that is optimized using coordinate descent in used implementation [143].

3.2.7 Elastic Net

Elastic Net (EN) represents an extension of Lasso regression. It extends the Lasso by adding another penalty term [222].

The problem can be written as [143, 222]:

$$\min_{\boldsymbol{w}} \frac{1}{2m} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}\|_2^2 + \alpha\rho\|\boldsymbol{w}\|_1 + \frac{\alpha(1-\rho)}{2}\|\boldsymbol{w}\|_2^2, \tag{3.27}$$

where α and L1 ratio ρ are parameters.

3.2.8 Bayesian Ridge regression

Bayesian Ridge regression tries to find maximum a-posteriori solution $\boldsymbol{w}$ to a probabilistic model where y is to have Normal distribution around $\boldsymbol{X}\boldsymbol{w}$ [143]:

$$\mathrm{p}\left(y|\boldsymbol{X}, \boldsymbol{w}, \alpha\right) = \mathrm{N}\left(y|\boldsymbol{X}\boldsymbol{w}, \alpha\right). \tag{3.28}$$

where α is from Gamma distribution with parameters α_k (shape) and α_θ (inverse scale). The prior over $\boldsymbol{w}$ is:

$$\mathrm{p}\left(w|\lambda\right) = \mathrm{N}\left(y|O, \lambda^{-1}\right). \tag{3.29}$$

where λ is from Gamma distribution with parameters λ_k (shape) and λ_θ (inverse scale).

3.2.9 Random Forest

Random Forests are popular ensemble estimators introduced in [22] that are very similar to more general methods *random subspaces* and *bagging* [21, 71, 103, 157, 158, 218]. A random forest estimates a number of Decision Trees — each of these DTs is estimated using a random sample taken with replacement from the *training set* and also split in each node is made using only a random subset of features [22, 157, 158]. RFs performs very well in many settings while being robust to noise and furthermore can be easily parallelized [22].

The hyperparameters of RFs are the same as for DTs but contain also another hyperparameter — *number of trees*. Higher number of trees in RFs usually leads to similar or higher accuracy but is computationally more expensive.

The advantage of RFs is that they also provide a measure of importance of individual variables and thus they can be used for feature selection as part of data pre–processing for other estimators [11, 55, 66, 67]. Another advantage of RFs is that they can deal with missing values implicitly [67].

One of the most popular extensions of RFs are *rotation forests* [156] that maps the random subset of features in each node to new space — usually using Principal component analysis (PCA) [156], non-parametric discriminant analysis (NDA), Random projections (RP), Sparse random projections [104] or with Independent component analysis (ICA) [45].

3.2.10 AdaBoost.R2

Another used ensemble estimator is AdaBoost.R2 [40] which is a modified regression version of the famous AdaBoost ensemble estimator [48]. It sequentially fits estimators and each subsequent estimator concentrates on the samples that were predicted with higher loss.

The used algorithm implemented in [143] slightly differs from [40] as it allows to used the weights directly in the fitted estimator and not only for weighted sampling of features:

1. start algorithm $t = 0$
2. To each training samples assign initial weight
$$w_i^t := 1, i = 1, 2, \ldots, m \tag{3.30}$$
3. fit estimator t to the weighted training set with weights w_i^t
4. compute prediction $\widehat{y}_i^t$ using the estimator t for each sample i
5. compute loss l_i for each training sample
$$l_i^t = \text{loss}\left(\left|\widehat{y}_i^t - y_i\right|\right) \tag{3.31}$$
6. calculate average loss $\overline{l^t}$
7. calculate confidence β^t for the estimator (low β^t means high confidence in estimator t)
$$\beta^t = \frac{\overline{l^t}}{1 - \overline{l^t}} \tag{3.32}$$
8. update weights of training samples
$$w_i^{t+1} := w_i^t \cdot \left(\beta^t\right)^{\left(1 - l_i^t\right)}, i = 1, 2, \ldots, m \tag{3.33}$$
9. $t = t + 1$ continue to step 3 while the average loss $\bar{l}^t < 0.5$

where loss(e) : $\mathbb{R} \rightarrow [0, 1]$. The work [40] suggests following losses:

1. *linear loss*

$$l_i^t = \frac{|\widehat{y}_i^t - y_i|}{D} \tag{3.34}$$

2. *square loss*

$$l_i^t = \frac{|\widehat{y}_i^t - y_i|^2}{D^2} \tag{3.35}$$

3. *exponential loss*

$$l_i^t = -\exp\left(\frac{-|\widehat{y}_i^t - y_i|}{D}\right) \tag{3.36}$$

where D is defined as

$$D = \sup\left\{\left|\widehat{y}_i^t - y_i\right|, i \in \{1, 2, \dots, m\}\right\} \tag{3.37}$$

3.2.11 Neural networks

Artificial Neural Networks (NNs) represent a very broad field of approaches and algorithms whose beginnings are in the 19th Century as the early NNs are based on the *linear regression* [161]. However, the early NNs differed significantly from the mpst commonly used NNs today as they were not able to learn from data [125, 161]. Some researchers consider the introduction of McCulloch's neural logical calculus [125] as the beginning of the field. Then appear NNs that were able to learn from data in unsupervised manner (1949) [69] (ref. from [161]) and only later in supervised manner, for example the perceptron algorithm in 1958 [159]. The popularity of NNs rises with the introduction of general learning algorithm — the *backpropagation*. The *backpropagation* was first introduced in 1970 [112, 161] but it was used in context of NNs a decade later — in 1981 [161] in [196].

A great number of various types of NNs was proposed since then — e.g. textitperceptron [159], *RBF* networs [23], *Boltzmann Machines* [70], *Hopfield* networks [77], *Restricted Boltzmann Machines* [173], *Deep Belief networks* (DBN) [15], *Recurrent neural networks* (RNNs) [44], *GRU networks* (networks with *gated recurrent units*) [27], *LSTM networks* (networks with *Long short-term memory* units) [72], *Auto–Encoders* (AE) [20], *Variational Auto–Encoders* (VAE) [94], *Sparse Auto–Encoders* (SAE) [152], *Denoising Auto–Encoders* [191], *convolutional neural networks* CNNs [49, 107], *Deconvolutional networks* [213], *Deep convolutional inverse graphics networks* (DCIGN) [101], *Generative Adversarial Networks* [59], *Neural Turing Machines* (NTM) [61, 211], *bidirectional recurrent neural networks* (BRNNs) [162], *Self-organizing maps* (SOM) [97], *Liquid State Machines* [120], *Extreme Learning Machines* (ELMs), *Recombinator networks* [74], and *Echo State networks* [82]. This list is incomplete and the types are not disjoint categories, the list was given only to show that the field of NNs is very broad and does not limit to the few simple architectures that are used in this work.

This section only provides a very brief overview of the relevant parts of the field focusing mostly on *feed forward* neural networks with simple architectures. For more detailed introduction into NNs, viz [60, 85].

3.2.11.1 Description

A NN can be defined as a computational graphical model with weighted edges. Nodes are called *neurons* and edges describes the flow of a signal. A neuron takes all input signals and combines them into a single output signal that is either part of the output of the NN or is sent to one or more other neurons. Thus for a neuron i with inputs $x_0, x_1, \ldots, x_n$ that came through edges with weights $w_0, w_1, \ldots, w_n$, an activation function $f(x)$ and a bias constant b the output x_i at time t is:

$$x_i[t] = f\left(b_i + \sum_{j=0}^{n} w_{i,j} x_j[t-1]\right) \tag{3.38}$$

The Eq. 3.38 is often written in a matrix form (without the time indices that are often unnecessary):

$$x_i = f\left(\boldsymbol{w}_i^{\mathrm{T}} \boldsymbol{x}\right), \tag{3.39}$$

where

$$\boldsymbol{x} = \begin{bmatrix} 1 \\ x_0 \\ \vdots \\ x_n \end{bmatrix} \tag{3.40}$$

and

$$\boldsymbol{w}_i = \begin{bmatrix} b_i \\ w_{i,0} \\ \vdots \\ w_{i,n} \end{bmatrix} \tag{3.41}$$

There are many possible activations functions (e.g. *linear* or *ReLU*) but the most known are the *sigmoid* activation functions and especially the *logistic function*:

$$f(x) = \frac{1}{1 + \mathrm{e}^{-x}} \tag{3.42}$$

An example of typical NN is shown in fig. 3.1 that contains a NN consisting of 3 layers (certain NNs can have more than several hundreds of layers — e.g. *residual networks* (ResNets) in [68]).

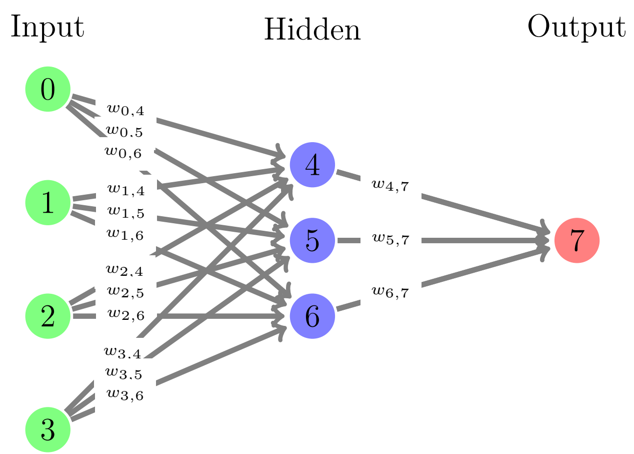


Figure 3.1: An example of a simple feedforward NN with 3 input neurons, 2 neurons in the hidden layer and 1 output neuron.

The output y of the NN from fig. 3.1 for given inputs $x_0, \ldots, x_3$ is:

$$y = \mathrm{S}\left(b_7 + w_{4,7}\mathrm{S}\left(b_4 + \sum_{j=0}^{3} w_{j,4}x_j\right) + w_{5,7}\mathrm{S}\left(b_5 + \sum_{j=0}^{3} w_{j,5}x_j\right) + w_{6,7}\mathrm{S}\left(b_6 + \sum_{j=0}^{3} w_{j,6}x_j\right)\right) \tag{3.43}$$

where S is the sigmoid activation function and $w_{i,j}$ represents the weight of the connection between neurons i and j. The computation is usually done by layers to avoid duplicate computations if the network has more layers (only the inputs $x_0, \ldots, x_3$ occur repeatedly in eq. (3.43)).

3.2.11.2 Training

The NNs are often trained using *gradient descent* with the use of *backpropagation* (BP) for computing the gradient of the NN with the respect to the used loss function. BP is basically an iterative application of the *chain rule of differentiation* [60]. Let $x \in \mathcal{R}$, $f(x) : \mathcal{R} \to \mathcal{R}$, $g(x) : \mathcal{R} \to \mathcal{R}$, $y = g(x)$, and $z = f(y) = f(g(x))$, then the chain rule of differentiation [60] is:

The NNs are usually written in terms of vector — for $\boldsymbol{x} \in \mathcal{R}^m$, $\boldsymbol{y} \in \mathcal{R}^n$, $z \in \mathcal{R}$, $g(\boldsymbol{x}) : \mathcal{R}^m \to \mathcal{R}^n$, $f(\boldsymbol{y}) : \mathcal{R}^n \to \mathcal{R}$, $\boldsymbol{y} = g(\boldsymbol{x})$, and $z = f(\boldsymbol{y})$, then the chain rule of differentiation [60] is:

$$\frac{\partial z}{\partial x_i} = \sum_{j=1}^{n} \frac{\partial z}{\partial y_j} \frac{\partial y_j}{\partial x_i} \tag{3.44}$$

Once there is a method for computing the gradient, then the gradient descent might be used for finding a local optima. Gradient descent works by going down in the loss function landscape [85]. The basic version of gradient descent make a fixed length step in the direction of highest gradient, however many other modifications to the algorithms were proposed in the literature — the gradient might be computed for only a small subset of the training samples at a time (*online gradient descent* or *minibatch gradient descent* — both are sometimes called *stochastic gradient descents* [85]), the algorithm might use momentum (or Nesterov momentum), or the algorithm might change the learning rate. Some of the popular algorithms for optimizing NNs besides the *stochastic gradient descent* are *Adadelta* [212], *adagrad* [42], *Adam* and *AdaMax*[93], *ESGD* [33], *Nadam* [39] or *RMSprop* [183]. More details about optimization of NNs is available in [60].

3.3 Forecast evaluation

There are many different measures that might be used for forecast evaluation — the most common ones are *mean absolute error* (MAE), *root mean square error* (RMSE) and *mean absolute percentage error* (MAPE). A more comprehensive list of available measures is in [81, 163]. This work uses MAE and RMSE; the MAPE is not used due to its unsuitable characteristics. The MAPE is not defined when the actual value is zero and explodes for values close to zero [81, 92, 184] and the MAPE is biased toward lower forecasts [81, 184].

3.3.1 Mean absolute error

The MAE is defined as:

$$\mathrm{MAE} = \frac{1}{n}\sum_{t=1}^{n} |y_t - \widehat{y_t}|, \tag{3.45}$$

where n is the size of the sample, y_t is the actual value of the time series at time t and $\widehat{y_t}$ is the predicted value for time t. It is the recommended error measures for most problems because it has clear interpretation and it is robust and not sensitive to outliers [10, 81].

3.3.2 Root mean square error

The RMSE is defined as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n} (y_t - \widehat{y_t})^2}, \tag{3.46}$$

where n is the size of the sample, y_t is the actual value of the time series at time t and $\widehat{y_t}$ is the predicted value for time t.

RMSE is one of the most popular error measures but as it puts larger penalty on larger errors, it is sensitive to outliers and is often misinterpreted as an average error [10, 81]. It is, however, very useful for most optimizations as it does not contain absolute value function and it is smooth and differentiable convex function and leads to stable solutions. Most of the used algorithms are, in fact, finding an estimator minimizing the RMSE, thus it the RMSE is used as one of main measures for evaluating the forecasts.

3.3.3 Diebold–Mariano test

The Diebold–Mariano (DM) test is often used for comparing the quality of two different forecasts [37]. This test allows for testing the difference of two forecasts with the null hypothesis of no difference. It allows for forecast errors to be non–Gaussian, with non–zero mean, serially correlated and contemporaneously correlated [37]. It uses the *loss differential* $L(e_t)$ for forecast error e_t where the function L depends on the used loss, e.g. quadratic loss is $L(e_t) = e_t^2$. Thus the time–t loss differential between forecasts A and B is $d_{ABt} = L(e_{At}) - L(e_{Bt})$ [36]. The only requirement of the DM test is that the loss differential is covariance stationary, i.e. the following three equations must hold:

$$\mathrm{E}(d_{ABt}) = \mu, \forall t, \tag{3.47}$$

$$\mathrm{Cov}(d_{ABt}, d_{AB(t-\tau)}) = \gamma(\tau), \forall t, \tag{3.48}$$

$$0 < \mathrm{Var}(d_{ABt}) = \sigma^2 < \infty. \tag{3.49}$$

The null hypothesis consist thus to $\mathrm{E}(d_{ABt}) = 0$ under which the DM statistics:

$$DM_{AB} = \frac{\overline{d_{AB}}}{\widehat{\sigma_{\bar{d_{AB}}}}} \rightarrow \mathrm{N}(0, 1), \tag{3.50}$$

where $\overline{d_{AB}} = \sum_{t=1}^{n} d_{ABt}$ is the sample mean of d_{ABt} and $\widehat{\sigma_{\overline{d_{AB}}}}$ is a consistent estimate of standard deviation of $\overline{d_{AB}}$ [36]. More details about the whole procedure are available in [36, 37].

3.3.4 Pareto rank

The estimators were evaluated using both RMSE and MAE which resulted in two rankings of estimators as the RMSE and MAE error measures are two competing objectives. The Pareto ranking was used to combine the two measures into a single ranking based on multi-objective optimization. A estimator is considered *Pareto efficient* when there is no other estimator that would be at least as good in one measure and strictly better in the other measure. The Pareto optimal set (frontier) is a set of all Pareto efficient estimators.

The Pareto rank is calculated iteratively — first, the Pareto optimal set $\mathbb{P}_1$ (frontier) is determined for the set $\mathbb{S}_1$ of all models and all models from the set $\mathbb{P}_1$ are assigned rank 1 and are removed from the set $\mathbb{S}_2 = \mathbb{S}_1 \setminus \mathbb{P}_1$, then another iteration is calculated.

The Pareto rank allows for comparing estimators without the need to limit the comparison to either RMSE or MAE measures, however, the model within a single Pareto rank might be very different while model with different Pareto ranks might still be very similar in terms of MAE and RMSE — the Pareto rank does not weight the individual RMSE and MAE measures. If one estimator had RMSE rank 1 while the MAE rank 100, the Pareto rank will still be 1. The final estimator should be selected from estimators with Pareto rank according to the preferences to the used error measures.

Chapter 4

Data

There are several important characteristics of the Electricity Markets such as occurrence of negative prices, intraday cycles and abrupt spikes. This chapter describes the price data on the Czech Day–ahead electricity markets and confirms that this market has all three characteristics. The data were obtained from the interval of 2009/1 — 2017/5 and consists of one Excel file per year. The data are provided by OTE, a.s. and are available at `http://www.ote-cr.cz/statistika/rocni-zprava`. The Excel files contains hourly marginal prices in euros per megawatt hour together with the amount of electricity sold and bought and the amount of electricity exported to or imported from Slovakia. It also contains several dependent variables that were calculated from the variables mentioned above — e.g. the total price for the electricity or daily averages. More detailed description of the published data is available at `http://www.ote-cr.cz/`. It is important to note, however, that the hourly prices in euros per megawatt hour were calculated from the price in CZK/MWh using the exchange rate provided by the Czech National Bank until 31.1.2009. Thus the prices until 31.1.2009 are not exactly the same as the prices after 31.1.2009 due to the slightly varying exchange rate but this difference is negligible for our purposes. The marginal prices in CZK/MWh since then are only informative and are calculated from the prices in EUR/MWh using the exchange rate provided by the Czech National Bank.

4.1 Price data

The data consists of 3073 sets of 24 hourly data from 2009/1 — 2017/5 a. The highest recorder price during that period reached 170 EUR/MWh and the lowest price reached -150 EUR/MWh, while the average electricity price is 38.42 EUR/MWh, the median is 37.44 EUR/MWhand the 25% and 75% quantiles were 28.00 EUR/MWh and 48.38 EUR/MWh — this is just an illustration of the extreme ranges of the price of electricity. The negative price of the electricity was observed in total of 284 samples, which is not very frequent as it represents less than 0.4 % of the dataset. However, even these rare negative prices might be very important for risk managers as the prices might be very low. The basic statistics for individual hours are shown in table 4.1 however, these static statistical description is only illustrative as there are many seasonalities and periodicities present in the electrical prices [195, 197].

hour	**count**	**mean**	**std**	**min**	$\mathbf{p_{25\%}}$	$\mathbf{p_{50\%}}$	$\mathbf{p_{75\%}}$	**max**
01	3103	29.317	10.997	-109.31	24.050	30.00	36.000	55.50
02	3103	26.291	11.134	-120.00	21.000	27.00	33.000	53.50
03	3103	24.454	11.304	-150.00	19.000	25.30	31.500	51.00
04	3103	22.929	11.358	-150.00	17.000	23.93	30.000	50.93
05	3103	23.543	11.355	-150.00	17.565	24.50	30.750	50.50
06	3103	26.873	11.607	-150.00	21.000	27.90	34.300	51.95
07	3103	34.371	14.597	-150.00	26.805	36.00	44.000	78.28
08	3103	42.092	18.061	-50.82	31.965	43.30	53.495	140.00
09	3103	44.654	17.498	-16.94	34.100	45.00	56.000	140.00
10	3103	45.428	16.243	-11.20	35.100	45.03	56.000	134.14
11	3103	44.879	15.764	-9.87	34.800	44.61	55.410	123.52
12	3103	44.913	16.000	-2.40	34.400	44.50	55.435	121.00
13	3103	43.073	15.836	-10.00	32.515	42.51	54.000	115.21
14	3103	41.150	15.876	-18.83	30.500	40.70	52.000	112.67
15	3103	39.530	15.344	-25.60	29.500	39.20	50.000	110.30
16	3103	39.457	15.002	-18.41	29.920	39.40	49.550	112.50
17	3103	40.843	15.286	-13.89	31.100	40.03	50.065	115.00
18	3103	45.241	17.441	-2.01	34.390	43.30	54.720	142.98
19	3103	48.785	17.490	-2.00	37.000	46.37	58.710	170.00
20	3103	49.075	15.442	-2.00	38.080	47.79	58.680	150.00
21	3103	46.168	12.885	-2.00	37.500	45.43	54.000	120.00
22	3103	40.959	10.637	-2.00	34.705	40.40	47.800	105.99
23	3103	38.580	10.080	-2.00	32.835	38.00	45.575	107.83
24	3103	32.480	9.773	-60.00	27.235	32.37	38.305	57.13

Table 4.1: The basic description of used price data for individual hours.

Another important characteristics beside the occurrence of negative prices is the occurrence of intraday cycles. The mean hourly prices showing the daily cycle is depicted in Fig. 4.1, however, as it contains only the average hourly price it does not tell us anything about the variance of the cycle. The Fig. 4.2 is just an illustration how the price might vary — this figure shows the hourly prices for all four quarters of the year 2016. The distribution of prices for individual hours also differs with the time of day as shown in fig. 4.3 and fig. 4.4. The fig. 4.3 shows the quite wide range of prices that occur at each hour in the dataset together with the few observation at the most extreme values. The boxplot in fig. 4.3 shows the range of the prices, the shape of distributions of hourly prices is better shown in fig. 4.4. Both plots show that the prices less vary during the night time where the demand is more stable — this tendecy occurs even when controlled for both yearly and monthly differences.

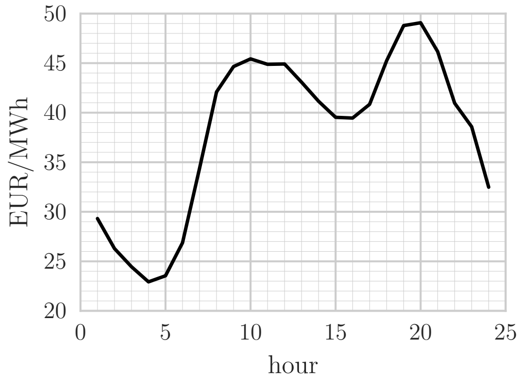


Figure 4.1: Mean hourly prices over the interval 2009/1-2017/5

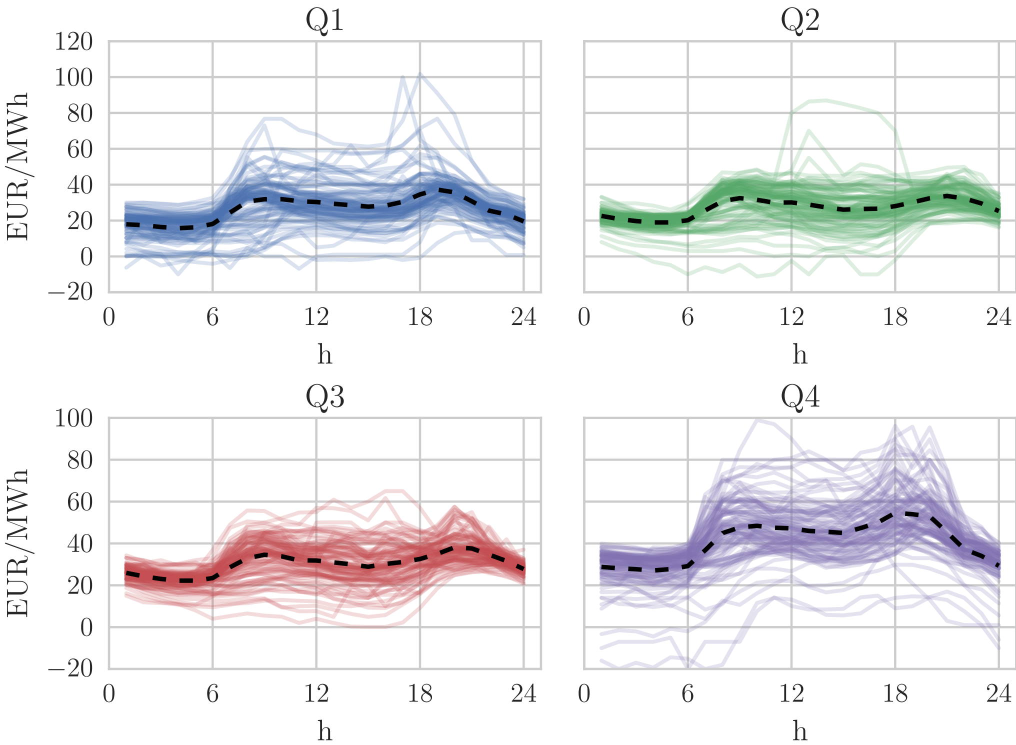


Figure 4.2: Daily prices for year 2016 for individual quarters.

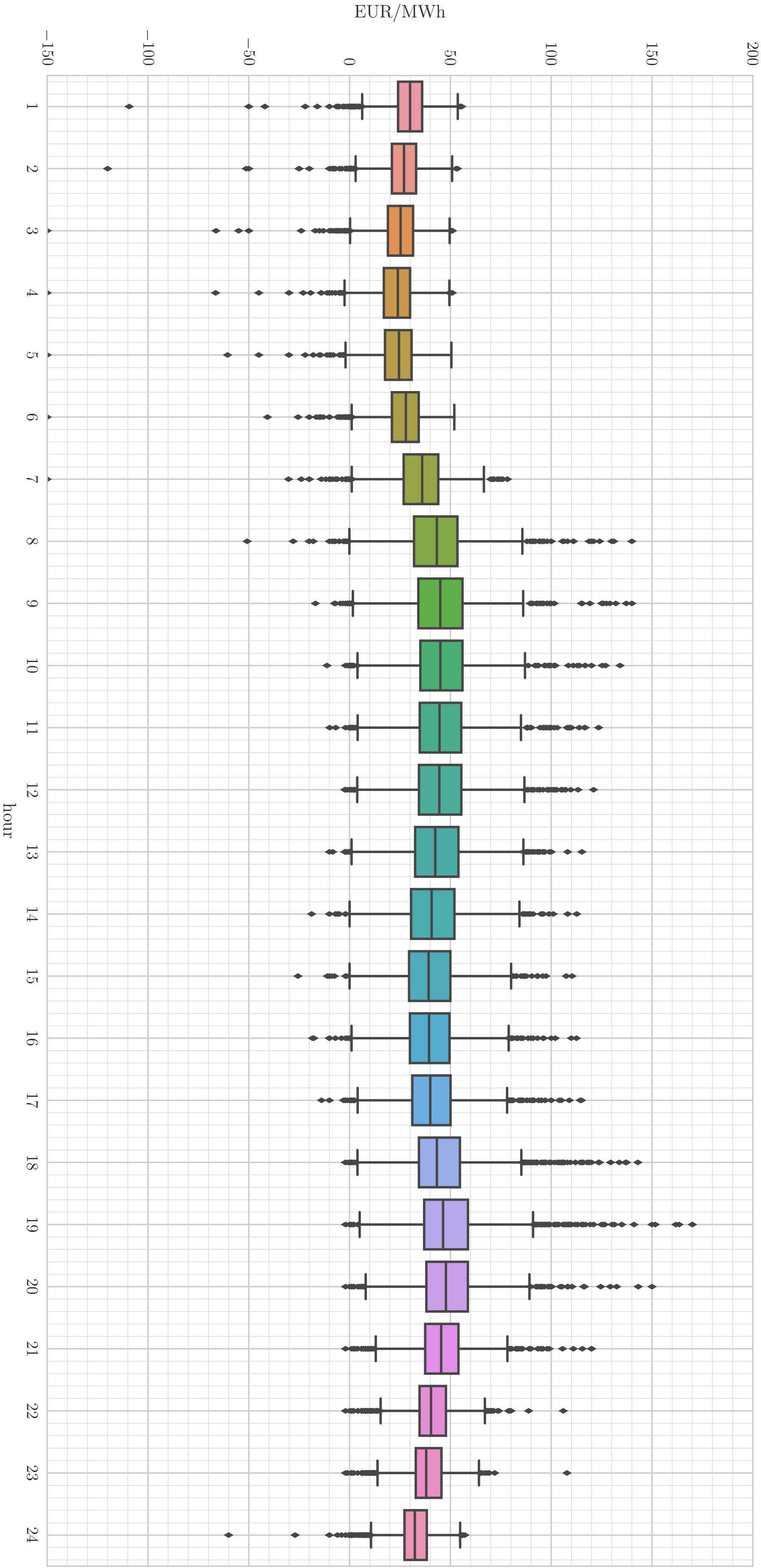


Figure 4.3: Boxplot of hourly prices for individual hours over the interval 2009/1-2017/5. The whiskers are at $1.5 \times IQR$ from the relevant quartile where IQR is the interquartile range.

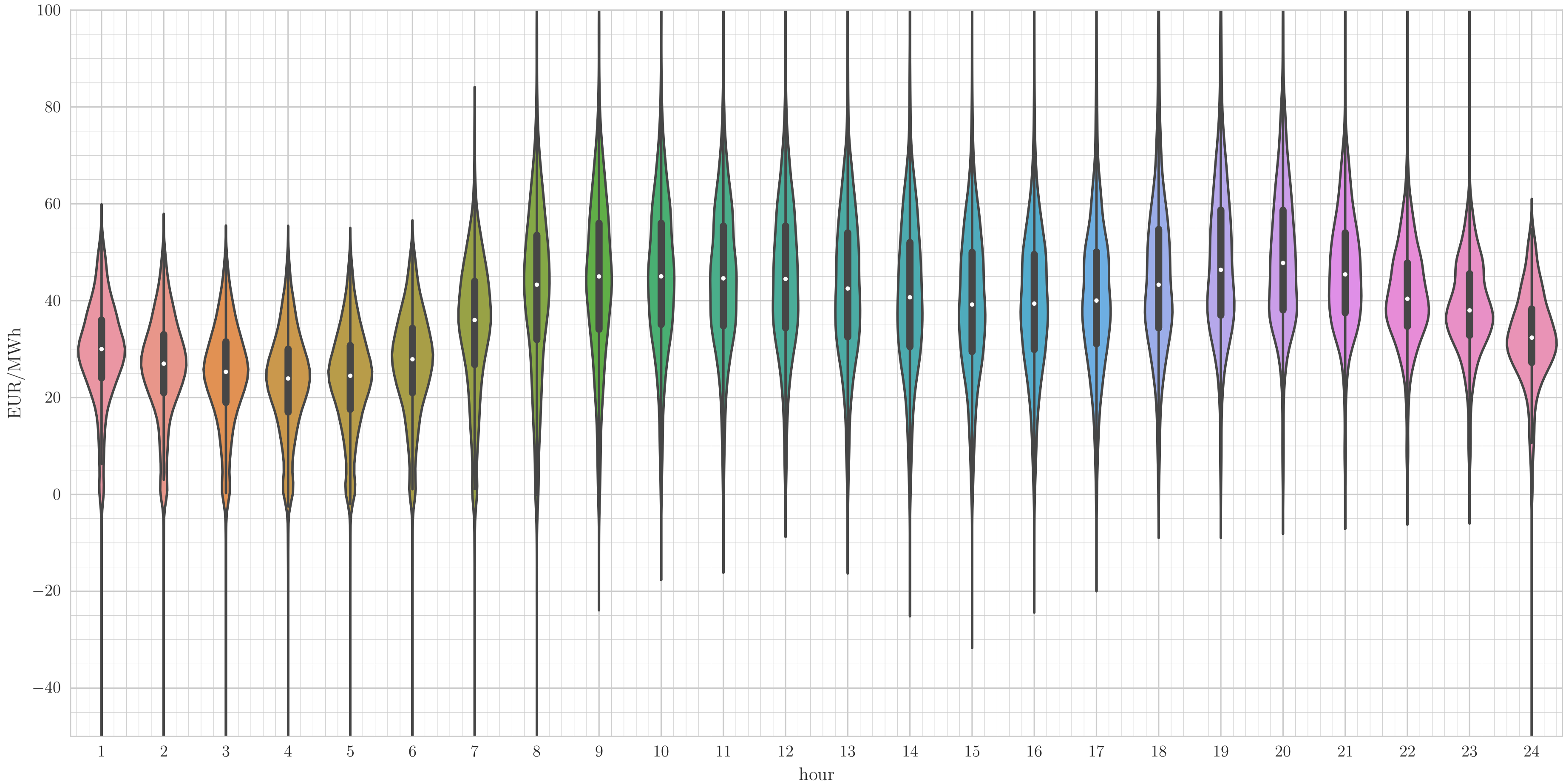


Figure 4.4: Distribution of hourly prices for individual hours over the interval 2009/1-2017/5. The quartiles are shown within each violin by black bars — the white dot depicts the median. The plot is cropped and some of the extreme values are not shown — viz fig. 4.3 for visualization of the distribution with outliers.

As can be seen from the Fig. 4.2, the price behaves slightly differently in each quarter which is partially caused by the weather prevailing in each season. The Fig. 4.7 shows the price averaged conditional on the quarter — we can observe that in the first and fourth quarter the second peak is much higher which is probably caused by the shorter period of daylight which causes the demand for electricity to go up as more electricity is needed for lightening. However the prices changes even in the long-term — the average hourly price for individual years is shown in Fig. 4.8. Furthermore, the differences between years are not only in the in the mean price but also the price distribution slightly varies between years as shown in fig. 4.5 which show the variation in yearly distribution of the price for hour slot for 19:00 – 20:00. The variations are even more visible when the data are shown for individual months as in fig. 4.6 which shows the variation in price for hour slot 11:00 – 12:00 for February. However, without more data, it is impossible to say whether the observed long-term change is due to the changes in the underlying energy industry or is caused by different weather in each year — most likely it is caused by both phenomena. The average electricity price for individual years and quarters is depicted in Fig. 4.9.

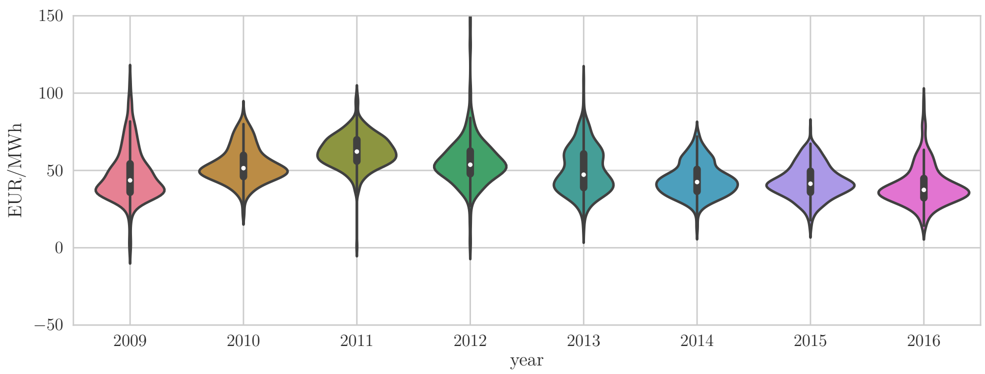


Figure 4.5: The variation in yearly distribution of the price for hour slot for 19:00 – 20:00 across all months.

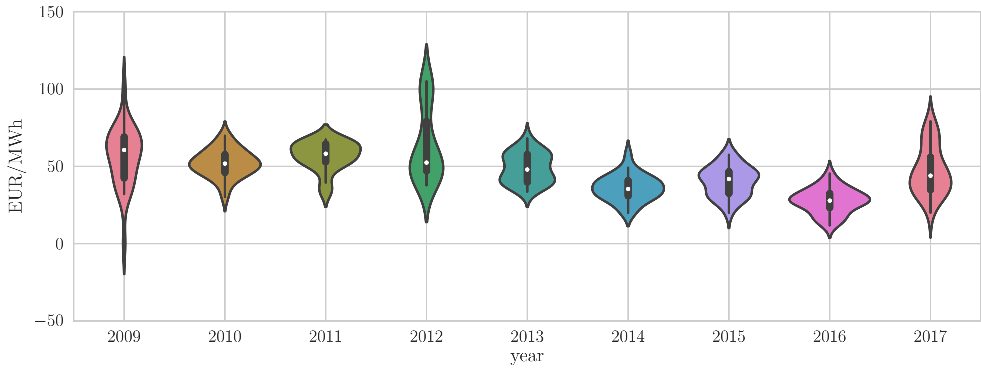


Figure 4.6: The variation in yearly distribution of the price for hour slot for 19:00 – 20:00 for February.

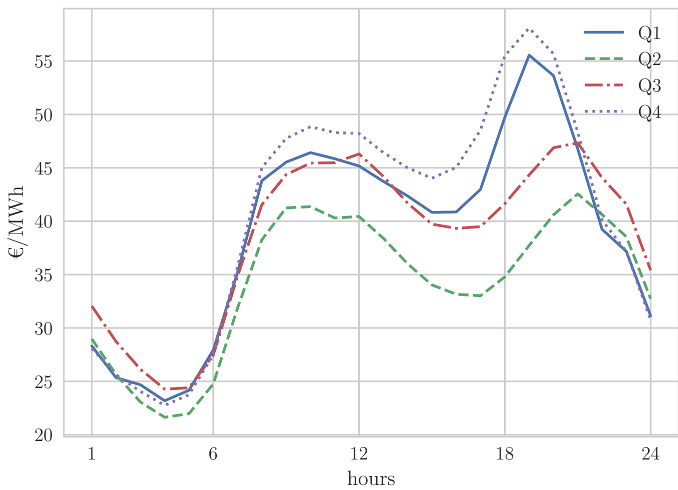


Figure 4.7: Mean price for individual quarters.

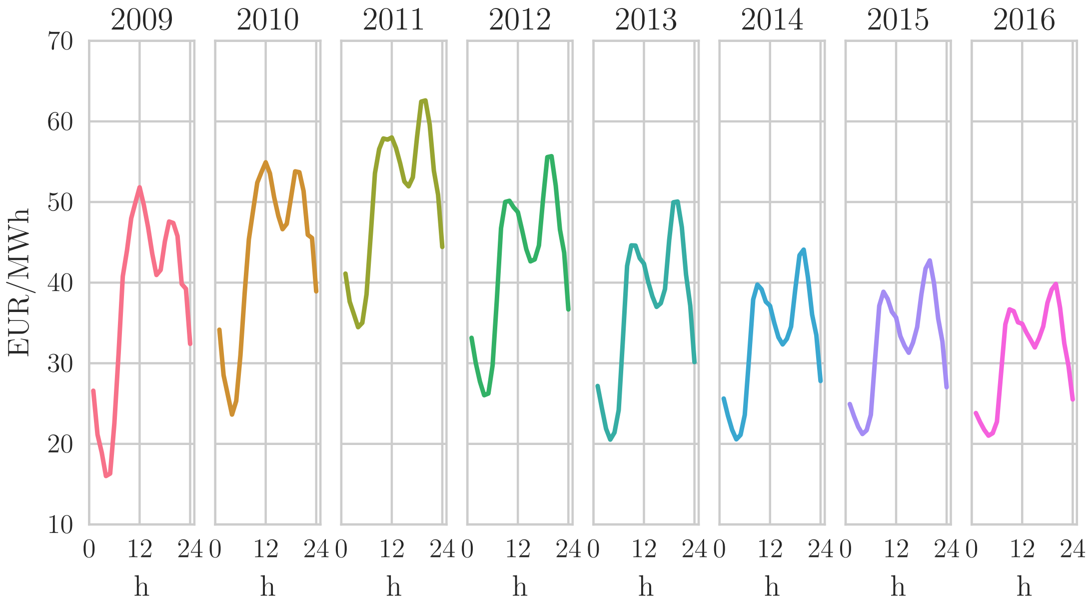


Figure 4.8: Mean price for individual years.

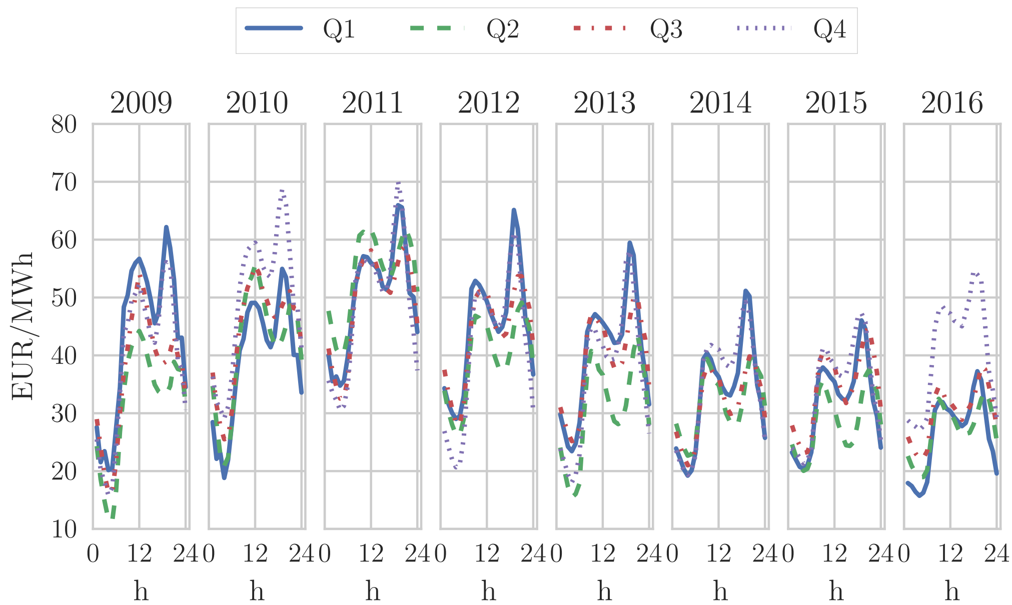


Figure 4.9: Mean price for individual years and quarters.

4.2 Weather data

As both demand and supply of electricity is influenced by weather [195, 197], this work also uses daily weather data for the electricity price prediction. The estimators used historical weather information — at time t the estimators have available weather at time $t-2$ or even older. While use of historical weather forecast from $t-2$ for the day t would be useful, no available open source of historical forecast has been found. The weather data were obtained from `https://www.wunderground.com`. The used variables were *temperature*, *dew point*, and *relative humidity*. Since all the variables have been measured many times within a day, a daily summary statistics were used, i.e. a single sample consists of recorded daily minimum, maximimum and average value for each of the three variables. The distrubition of the daily averages of the three variables are shown in fig. 4.10, fig. 4.11, and fig. 4.12 respectivelly.

4.3 Dataset creation

4.3.1 Division of data

The available data was divided into three disjoint datasets — *training* set, *validation* set and *test set*. The *training* dataset was used for training during parameter optimization for individual estimators, the performance of the parametrization were evaluated on the out-of-sample *validation data*. Once the optimal parametrization was found, the estimator was retrained using both *training* and *validation* data and its performance was evaluated using the *test data* that were previously unseen. The *validation* and *testing* periods were distributed in the whole time range — 4 months of *training* data were followed by 1 month of *validation data* and 1 month of *test* data, then 1 month of data was skipped to ensure the independency of *training* and *test* data as the highest lag that was used throughout this work is 31. Then the cycle repeated. Over the years 2010 to 2016, this resulted in 12 months of *test* data where each month was exactly once. The division of data is shown in table 4.2.

	1	2	3	4	5	6	7	8	9	10	11	12
2010	X	Tr	Tr	Tr	Tr	V	Te	X	Tr	Tr	Tr	Tr
2011	V	Te	X	Tr	Tr	Tr	Tr	V	Te	X	Tr	Tr
2012	Tr	Tr	V	Te	X	Tr	Tr	Tr	Tr	V	Te	X
2013	Tr	Tr	Tr	Tr	V	Te	X	Tr	Tr	Tr	Tr	V
2014	Te	X	Tr	Tr	Tr	Tr	V	Te	X	Tr	Tr	Tr
2015	Tr	V	Te	X	Tr	Tr	Tr	Tr	V	Te	X	Tr
2016	Tr	Tr	Tr	V	Te	X	Tr	Tr	Tr	Tr	V	Te

Table 4.2: The division of data into the *training* set, *validation* set and *test set*. The month denoted by `X` is skipped, by `Tr` put into the *training* set, by `V` put into the *validation* set and by `Te` put into the *test* set.

4.3.2 Scaling the data

Prior to training the individual estimators, the data were normalized, i.e. they were scaled and shifted in a such way that they had zero mean and standard deviation

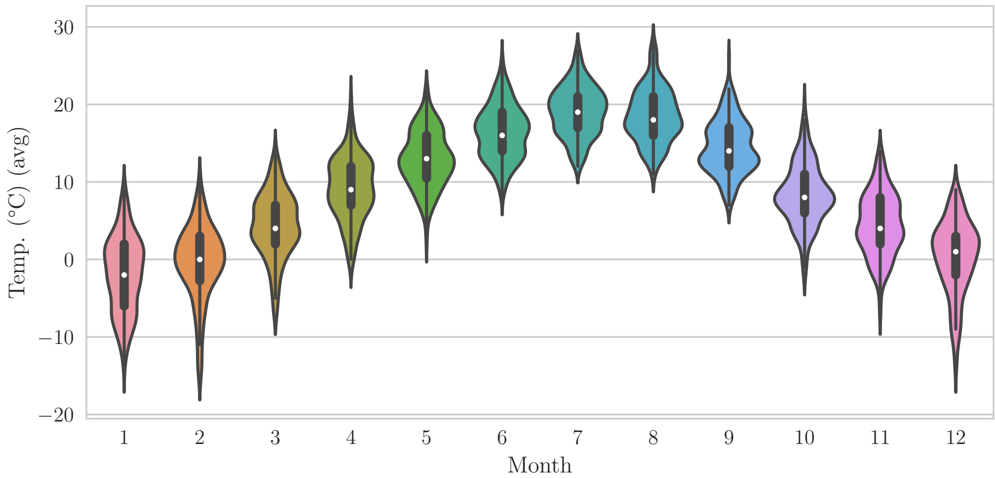


Figure 4.10: The distribution of average daily temperature for individual months.

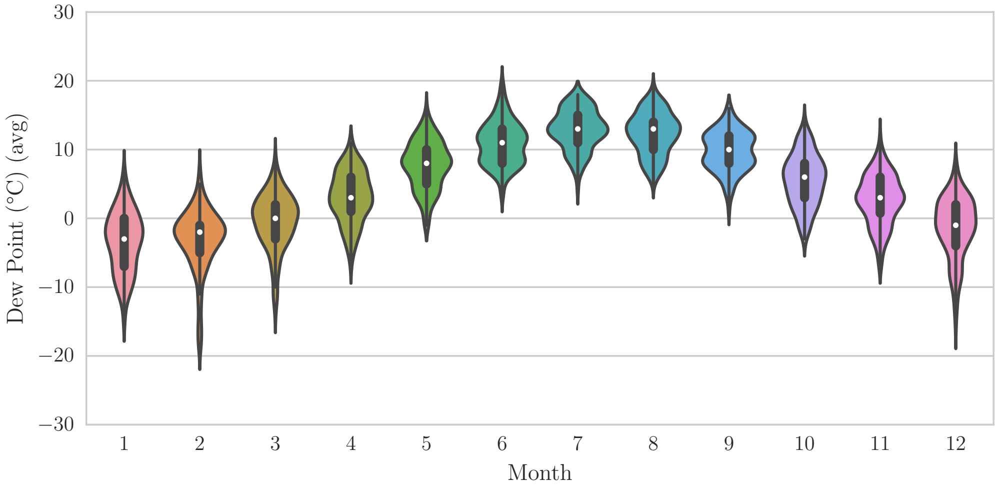


Figure 4.11: The distribution of average daily dew point for individual months.

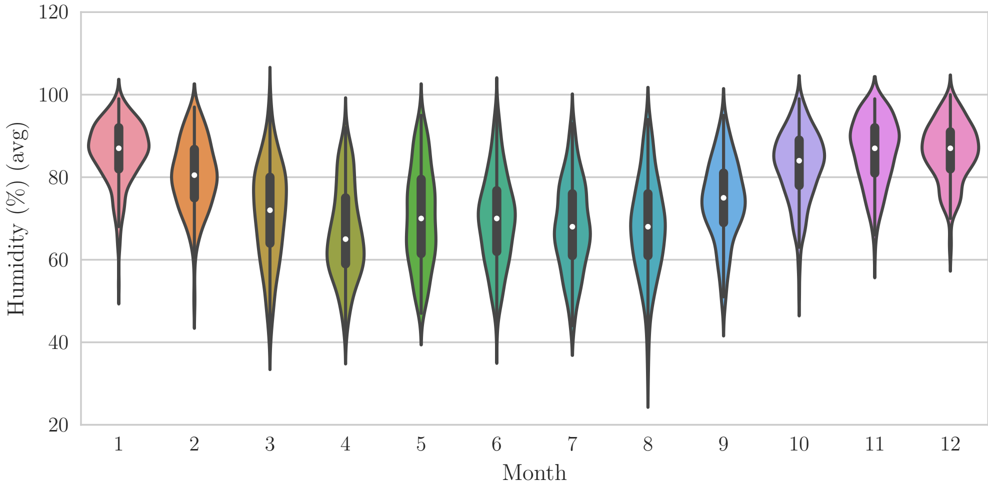


Figure 4.12: The distribution of average daily humidity for individual months.

equal to one over the *training* data. The parameters of the transform were estimated using the *training* data only and then used also for both *validation* and *test* data. Each forecast was then transformed back in order to evaluate the prediction in the actual units.

Chapter 5

Results

As several different classes and parametrizations of estimators were used, the results are analysed both from overall point of view and also within each class of estimators. The main measures used for comparing estimators were the RMSE and MAE on the out–of–sample test data. The estimators were ranked by their RMSE and MAE performance both in total and within the estimator's class only. Then 10 best estimators for both measures were selected for each class of estimators for further analysis. The 10 estimators' forecasts were then compared against each other using the Diebold–Mariano (DM) test to analyze the gap in quality between the top forecasts. The used significance level for the Diebold–Mariano test was 5 %, the power p of the DM test is used with the respect to the error measure used for selecting the estimators — power $p = 1$ is used for estimators selected with respect to MAE while power $p = 2$ (sqaured errors) is used for estimators selected with respect to RMSE.

For further analysis of the models from the overall comparison, only small subset is selected as the total number of used estimators (> 5000) makes comparing all models infeasible.

5.1 Comparison by class

The models were compared by the underlying estimator class. The models from within the class might still differ from each other by used parametrization as the individual models had optimized parameters with the respect to the used variables for most of the estimator classes. The methods used for optimization are described in section 3.1 while the actual optimization is described individually for each estimator class.

5.1.1 Used models

The individual estimators have assigned a small code to allow for easy identification of individual parametrizations. The name consists of several field — `[estimator class]_[used lag and mean variables]_[price only | weather data]-[used dummy variables]`.

The possible estimator class codes are

AB-DT AdaBoost with Decision Trees as the underlying weak regressor

AB-LR AdaBoost with Linear Regression (OLS) as the underlying weak regressor

ANN48-0.5d-linear ANN with 48 neurons in hidden layer and 0.5 dropout with tanh activation function in hidden layer and linear activation function in the output layer trained using the Nadam optimizer for 600 epochs and batch size 64. The data were shuffled during training. The hyperparameters of the Nadam optimizer [39]: initial learning rate $l = 0.002$, $v = 0.999$,$\mu = 0.9$, $\epsilon = 1 \times 10^{-8}$, and schedule decay of 0.004. The parameters were set as recommended in [26] (the parameters have different names: $v = \beta_2$ and $\mu = \beta_1$ in [26]).

ANN84-0.5d-1200e ANN with 84 neurons in hidden layer and 0.5 dropout with tanh activation function in hidden layer and linear activation function in the output layer trained using the SGD optimizer for 1200 epochs and batch size 64. The weights in the hidden layer were regularized using $L2$ regularizer with weight 0.2. The data were shuffled during training. The hyperparameters of the SGD optimizer: learning rate $l = 0.01$, momentum $m = 0$ and decay $d = 0$, i.e. vanialla SGD was used.

BR Bayesian Ridge regression. Used parameters werre: maximum number of iteration during optimization $n = 300$, tolerance $t = 0.001$, the parameters α_1, α_2, λ_1, and λ_2 were optimized individually for each estimator.

EN Elastic Net regression, both its parameters α and L1 ratio r were optimized individually for each estimator.

KRR-linear Kernel Ridge Regression with linear kernel (equiv. to Ridge Regression). The parameter α was optimized individually.

KRR-poly-2 Kernel Ridge Regression with polynomial kernel with degree $d = 2$, while the parameter γ was set to $\gamma = \frac{1}{\text{number of samples}}$ as default in [143].

Lars Least Angle Regression (LARS) where the number of non-zero coefficients was optimized using grid search over several possible values.

Lasso Lasso Regression (numerical optimization) where the maximum number of iteration during optimization was set to $n = 5000$. The parameter α was optimized individually.

LassoLars Lasso Regression with exact solution based on modified LARS viz section 3.2.6.1. The parameter α was optimized individually.

OLS Linear regression using Ordinary Least Squares (OLS).

RF Random forest regression with 2000 trees in the ensemble. The maximum depth was set to 20, the maximum features at each node was set to $\log_2(\text{number of features})$ and the used criterion for optimization was *mean squared error*.

SVR-linear-e Suppor Vector Regression with linear kernel and L_1 loss (ϵ-*insensitive* loss). The tolerance parameter was set to 0.0001 and the maximum number of iterations was set to 1000 as default in [143]. Parameters ϵ and C optimized individually.

SVR-linear-e2 Suppor Vector Regression with linear kernel and L_2 loss (*squared ϵ-insensitive* loss). The tolerance parameter was set to 0.0001 and the maximum number of iterations was set to 1000 as default in [143]. Parameters ϵ and C optimized individually.

SVR-poly-2 Suppor Vector Regression with polynomial kernel with degree $d = 2$ and L_1 loss (*ϵ-insensitive* loss). The tolerance parameter was set to 0.001 and the maximum number of iterations was not limited, the coeffient of the polynomial kernel $c_0 = 0$ (*homogeneous polynomial kernel*) as default in [143]. Parameters ϵ and C optimized individually.

SVR-poly-3 Suppor Vector Regression with polynomial kernel with degree $d = 3$ and L_1 loss (*ϵ-insensitive* loss). The tolerance parameter was set to 0.001 and the maximum number of iterations was not limited, the coeffient of the polynomial kernel $c_0 = 0$ (*homogeneous polynomial kernel*) as default in [143]. Parameters ϵ and C optimized individually.

SVR-poly-4 Suppor Vector Regression with polynomial kernel with degree $d = 4$ and L_1 loss (*ϵ-insensitive* loss). The tolerance parameter was set to 0.001 and the maximum number of iterations was not limited, the coeffient of the polynomial kernel $c_0 = 0$ (*homogeneous polynomial kernel*) as default in [143]. Parameters ϵ and C optimized individually.

SVR-rbf Suppor Vector Regression with RBF kernel and L_1 loss (*ϵ-insensitive* loss). The tolerance parameter was set to 0.001 and the maximum number of iterations was not limited as default in [143]. Parameters ϵ and C optimized individually.

SVR-sigmoid Suppor Vector Regression with sigmoid kernel and L_1 loss (*ϵ-insensitive* loss). The tolerance parameter was set to 0.001 and the maximum number of iterations was not limited, the coeffient of the sigmoid kernel $c_0 = 0$ (*homogeneous sigmoid kernel*) as default in [143]. Parameters ϵ and C optimized individually.

The possible lag and mean variables codes are

T2 only the data from day $t - 2$ are included.

T2-7 the data from days $t - 2$ to $t - 7$ are included.

T2-31 the data from days $t - 2$ to $T - 31$ are included.

T2,3,7 the data from days $t - 2$, $t - 3$, and $t - 7$ are included.

T2,7 the data from days $t - 2$ and $t - 7$ are included.

T2,7,14,28 the data from days $t - 2$, $t - 7$, $t - 14$, and $t - 28$ are included.

T2,M7 the data from day $t-2$ and the mean of hourly data for days $t-2$ to $t-7$ are included.

T2,M7,30 the data from day $t - 2$ and the means of hourly data for days from $t - 2$ to $t - 7$ and for days from $t - 2$ to $t - 30$ are included.

T2,7,M7 the data from days $t-2$ and $t-7$ and the mean of hourly data for days $t - 2$ to $t - 7$ are included.

T2,7,14,28,M7,30 the data from day $t-2$, $t-7$, $t-14$, and $t-28$ and the means of hourly data for days from $t-2$ to $t-7$ and for days from $t-2$ to $t-30$ are included.

nl no historical data included.

If only prices are included, **p** is appended to the name of the estimator; if weather data are includes, **w** is appended. Possible codes for dummies:

no code no dummy variable used.

W Weekend dummy variable included. Cannot be used with dummy D.

D a dummy variable included for days from Tuesday to Sunday, Monday is the base day. Cannot be used with dummy W.

M a dummy variable included for months from February to December, January is the base month.

Y a dummy variable included for year from 2011 to 2016, year 2010 is the base year.

An estimator is created for each possible combination of estimator class, used lag and mean variables, price only or weather data , and the dummy variables, which resulted in 5230 different estimators.

5.1.2 AdaBoost with Decision Trees (*AB-DT*)

This estimator class consists of AdaBoost estimators with Decision Trees as the underlying weak regressor. All estimators from this class consist of 24 sub-estimators, each predicting a single hour.

5.1.2.1 Optimization

The parameter *learning rate* was optimized for each of the estimator using the NM method, the starting point for optimization was found using the PSO for model `AB-DT_T2,7,14,28,M7,30_w-WMY`. The PSO used 10 particles for 5 generations as the optimization was quite costly and the parameter space was $lr \in [0, 0.3]$. The starting point for NM was learning rate $lr \approx 0.0500$.

5.1.2.2 Best estimators

The 10 best estimators were selected by both MAE and RMSE as shown in table C.1 and table C.2 for further analysis using the DM test. The best estimators by Pareto rank are shown in table 5.1 which also contains most of the estimators selected by both measures.

The results of the DM test at 5% significance level are shown in fig. C.1a. While there are significant differences in the forecasts betwen the estimators selected by MAE, this cannot be said for estimators selected by RMSE. This might have been caused by the `AB-DT` optimizing the RMSE error which made the estimators much more closely stacked in the RMSE measure compared to the MAE measures — the difference in RMSE between the 1st estimator and 10th by RMSE is 0.0315 EUR/MWh, while the difference in MAE between the 1st estimator and 10th by MAE is 0.0537 EUR/MWh. The 1st estimator is better by DM test then only the 10th estimator by RMSE at 5% significance level. This estimator class prefered the `T2,7,14,28,M7,30` variables with weekend dummy variables for both the measures. The weather information was not very useful for this class of estimators as it is not very frequent in the selection of the best estimators. The inclusion of weather variables had small negative effect as is visible in figs. A.1 and A.2, which might have been caused by the tendency of AdaBoost estimators to overfit the data. While the overfiting for `AB-DT` is lower compared to `AB-LR` (and `OLS`, `RF`, `SVR-poly2`, `SVR-poly-3`) as can be seen in fig. B.1 and fig. B.2, it is present.

Overall, the `AB-DT` is not very successful estimator as its error is quite high compared to other estimators (viz figs. A.1 to A.8). It ranked 16th by both RMSE and MAE when compared to other estimators as shown in figs. 5.1 and 5.2, furthermore, its best estimator by RMSE is 2741st in overall RMSE rankings and the best by MAE is 2711th in overall MAE rankings.

The example prediction on test data is shown in fig. E.1b (best by RMSE) and fig. E.1a (best by MAE) for 14th – 28th May 2016.

rPareto	**Var.**	**W.**	**RMSE**	**MAE**	**rRMSE**	**rMAE**
1/264	T2,7,14,28,M7,M30-W	No	8.6204	6.3587	1/2741	3/2716
1/264	T2,7,14,28,M7,M30-WM	No	8.6234	6.3443	2/2744	1/2711
2/265	T2,7,14,28,M7,M30-WY	No	8.6258	6.3497	3/2747	2/2713
3/268	T2,M7,30-WMY	No	8.6372	6.3880	5/2759	6/2748
3/274	T2,7,M7-WY	No	8.6317	6.5221	4/2756	52/2852
3/267	T2,7,14,28,M7,M30-WMY	No	8.6572	6.3752	12/2775	4/2735
4/275	T2,M7-WY	No	8.6417	6.5421	6/2761	64/2865
4/269	T2,M7,30-W	No	8.6447	6.3980	7/2764	10/2762
4/269	T2,M7,30-WM	No	8.6506	6.3935	9/2772	8/2756
4/270	T2,7,14,28,M7,M30-WY	Yes	8.7161	6.3789	17/2795	5/2738
5/270	T2,M7,30-WY	No	8.6543	6.4125	11/2774	15/2771
5/275	T2,7,M7-WMY	No	8.6477	6.5339	8/2767	59/2859
5/271	T2,7,14,28,M7,M30-W	Yes	8.7190	6.3915	18/2798	7/2752

Table 5.1: The list of best AB-DT estimators with Pareto rank at most 5. Column *rPareto* contain the Pareto rank within the class and also the overal Pareto rank (equivalety for *rRMSE* and *rMAE*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

5.1.3 AdaBoost with Linear Regression (*AB-LR*)

This estimator class consists of AdaBoost estimators with Linear Regression (OLS) as the underlying weak regressor. All estimators from this class consist of 24 sub-estimators, each predicting a single hour.

5.1.3.1 Optimization

The parameter *learning rate* was optimized for each of the estimator using the NM method, the starting point for optimization was found using the PSO for model `AB-LR_T2,7,14,28,M7,30_w-WMY`. The PSO used 20 particles for 50 generations as the optimization was quite costly and the parameter space was $lr \in [0, 1]$. The starting point for NM was learning rate $lr \approx 0.000145$ — the prefered learning rate was converging to zero thus the model was quite close to OLS.

5.1.3.2 Best estimators

The 10 best estimators were selected by both MAE and RMSE as shown in table C.3 and table C.4 for further analysis using the DM test. The best estimators by Pareto rank are shown in table 5.2 which also contains most of the estimators selected by both measures.

The results of the DM test at 5% significance level are shown in fig. C.1b. There are signficant differences between the 1st estimator and the rest for both MAE and RMSE, there are also significant differences between the first three estimators and the last four estimators by RMSE. This estimator class prefered the `T2,M7,30` variables with weekend dummy variables for both the measures. Similarly as for `AB-DT`, the weather information was not very useful for this class of estimators as it is not very frequent in the selection of the best estimators. The inclusion of weather variables had small negative effect as is visible in figs. A.1 and A.2. This estimator was heavily overfitting to the training data as can be seen in fig. B.1 and fig. B.2. It is overfitting even more than the OLS to which it is similar when the learning rates are low.

Overall, the `AB-DT` is quite successful (it was the 6th best estimator by RMSE and 8th by MAE —- viz figs. 5.1 and 5.2) with similar scores to other top–performing estimator classes (viz figs. A.1 to A.8). Its best estimator by RMSE was the 157th in the overall ranking. The best estimator by MAE had slightly worse position — it was 252nd.

The example prediction on test data is shown in fig. E.2b (best by RMSE) and fig. E.2a (best by MAE) for 14th – 28th May 2016.

5.1.4 ANN with 48 neurons (*ANN48-0.5d-linear*)

This estimator class consists of ANN with 48 neurons in the hidden layer with tanh activation function and 0.5 dropout.

5.1.4.1 Optimization

This estimator class was not optimized on individual estimator basis but the suitable number of neurons and the intensity of dropout was determined by experiments using PSO and grid search.

rPareto	Var.	W.	RMSE	MAE	rRMSE	rMAE
1/38	T2,M7,30-D	No	7.4500	5.4696	1/157	1/252
2/45	T2,M7-D	No	7.4969	5.5392	2/222	2/468
3/47	T2,7,M7-D	No	7.4972	5.5454	3/224	3/497
4/62	T2-DMY	No	7.5834	5.5466	12/391	4/503
4/48	T2,7,14,28,M7,M30-D	No	7.5081	5.5539	4/238	9/540
5/64	T2,3,7-DMY	No	7.5848	5.5516	13/396	5/530
5/56	T2,M7,30-DM	No	7.5451	5.5735	6/299	11/619
5/56	T2,7,M7-DY	No	7.5449	5.5841	5/298	15/669

Table 5.2: The list of best AB-LR estimators with Pareto rank at most 5. Column *rPareto* contain the Pareto rank within the class and also the overal Pareto rank (equivalety for *rRMSE* and *rMAE*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

5.1.4.2 Best estimators

The 10 best estimators were selected by both MAE and RMSE as shown in table C.6 and table C.5 for further analysis using the DM test. The best estimators by Pareto rank are shown in table 5.3 which also contains most of the estimators selected by both measures.

The results of the DM test at 5% significance level are shown in fig. C.1c. While the inclusion of weather data had a small negative effect on average as shown in-figs. A.1 and A.2, the top 10 selection contain several estimators with weather data. This estimator was also heavily overfitting to the training data as can be seen in fig. B.1 and fig. B.2 though not as extremely as `AB-LR` and `OLS` estimators.

Overall, the `ANN48-0.5d-linear` was in middle of rankings when compared the best estimators from each class, however, it was better than the similar ANN but with 84 neurons and longer training when comparing forecasts selected by both RMSE and MAE as shown in figs. 5.1 and 5.2. This estimator class ranked 9th by MAE and 11th by RMSE, its best estimator by RMSE is 392nd in overall RMSE rankings and the best by MAE is 308th in overall MAE rankings.

The example prediction on test data is shown in fig. E.3b (best by RMSE) and fig. E.3a (best by MAE) for 14th – 28th May 2016.

5.1.5 ANN with 84 neurons (*ANN84-0.5d-1200e*)

This estimator class consists of ANN with 48 neurons in the hidden layer with tanh activation function and 0.5 dropout trained for 1200 epochs.

5.1.5.1 Optimization

This estimator class was not optimized on individual estimator basis but the suitable number of neurons and the intensity of dropout was determined by experiments using PSO and grid search.

rPareto	Var.	W.	RMSE	MAE	rRMSE	rMAE
1/63	T2,M7-D	No	7.5839	5.5672	1/392	8/599
1/52	T2,M7,30-D	Yes	7.5874	5.4947	2/404	1/308
2/66	T2,M7-DY	No	7.5895	5.5604	3/411	7/567
2/59	T2,M7,30-DY	No	7.6020	5.5170	4/454	2/367
3/69	T2-DMY	Yes	7.6082	5.5564	5/479	5/550
3/70	T2,M7-D	Yes	7.6186	5.5483	7/522	4/512
3/67	T2,7,14,28,M7,M30-D	Yes	7.6542	5.5313	15/640	3/431
4/72	T2,7,M7-D	No	7.6109	5.5822	6/492	14/660
4/74	T2,7,14,28,M7,M30-D	No	7.6289	5.5583	8/552	6/556
5/77	T2,7-DMY	No	7.6360	5.5751	9/574	11/627
5/77	T2,M7,30-D	No	7.6480	5.5685	11/615	9/602

Table 5.3: The list of best ANN48-0.5d-linear estimators with Pareto rank at most 5. Column *rPareto* contain the Pareto rank within the class and also the overal Pareto rank (equivalety for *rRMSE* and *rMAE*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

5.1.5.2 Best estimators

The 10 best estimators were selected by both MAE and RMSE as shown in table C.8 and table C.7 for further analysis using the DM test. The best estimators by Pareto rank are shown in table 5.4 which also contains most of the estimators selected by both measures.

The results of the DM test at 5% significance level are shown in fig. C.1d. When comparing the estimators selected by MAE, there are significant differentces at the 5% level betwen the three best and the 6th to 10th estimators, however, when comparing by RMSE, the differences are not significant for most of the estimators. This migh be again caused by thorough optimizing of the RMSE during training and thus the best estimators are closer to each other when comparing by RMSE then when comparing by MAE. The differences in between the 10th and 1st estimators seem to support this hypothesis: the difference between the 10th and 1st estimator by RMSE is 0.0598 while the difference for MAE is 0.0772.

While the inclusion of weather data had a almost no effect on average for MAE while having small negative average effect for RMSE as shown infigs. A.1 and A.2 — this is supported also by the selection of top 10 where a half of estimators selected by MAE uses the weather data while only two of estimators selected by RMSE.

This estimator was also slighty overfitting to the training data as can be seen in fig. B.1 and fig. B.2 though less than the network with only 48 neurons (usually, the more complex networks have higher capacity for overfitting, the fact that this more complex network was overfitting less might have been caused by high dropout with longer training period, however, further analysis is needed). Despite the lower overfitting compared to the less complex network, the `ANN84-0.5d-1200e` had slightly higher error.

Overall, the `ANN84-0.5d-1200e` is in the lower half when comparing best estimators' forecasts selected by both RMSE and MAE as shown in figs. 5.1 and 5.2 — it ranked 13th by both MAE and RMSE, its best estimator by RMSE is 677th in overall RMSE rankings and the best by MAE is 563rd in overall MAE rankings.

The example prediction on test data is shown in fig. E.4b (best by RMSE) and fig. E.4a (best by MAE) for 14th – 28th May 2016.

rPareto	Var.	W.	RMSE	MAE	rRMSE	rMAE
1/90	T3-31-DM	No	7.6662	5.6894	1/677	25/1230
1/82	T2,M7,30-DMY	Yes	7.7115	5.5598	4/832	1/563
1/92	T2,M7,30-D	No	7.6979	5.6188	2/790	9/820
2/86	T2-7-DM	Yes	7.7241	5.5664	8/887	2/595
2/92	T2,M7,30-D	Yes	7.7199	5.6070	6/868	6/765
2/99	T2,7,14,28,M7,M30-DMY	No	7.7107	5.6548	3/828	17/1034
3/100	T2,7,14,28,M7,M30-DM	No	7.7140	5.6848	5/843	22/1201
3/93	T2,M7,30-DY	Yes	7.7474	5.5963	13/987	4/717
3/99	T2,M7,30-DY	No	7.7235	5.6364	7/884	14/929
3/89	T2,7,14,28,M7,M30-DMY	Yes	7.7301	5.5996	9/917	5/730
3/91	T2-7-WMY	Yes	7.7676	5.5767	19/1076	3/635
4/102	T2-7-DY	Yes	7.8484	5.6168	54/1371	7/814
4/107	T2,M7,30-DM	No	7.7314	5.6796	10/923	21/1173
4/101	T2-7-DM	No	7.7433	5.6305	12/971	11/893
4/107	T2-7-WM	Yes	7.8020	5.6253	28/1210	10/859
4/102	T2-7-D	Yes	7.8088	5.6169	33/1237	8/817
4/102	T2-7-D	No	7.7342	5.6576	11/937	18/1052
5/113	T2,M7-DMY	No	7.7576	5.6876	14/1040	24/1221
5/105	T2-7-WMY	No	7.7699	5.6355	20/1089	13/925
5/105	T2-7-DY	No	7.7715	5.6335	21/1094	12/912

Table 5.4: The list of best ANN84-0.5d-1200e estimators with Pareto rank at most 5. Column *rPareto* contain the Pareto rank within the class and also the overal Pareto rank (equivalety for *rRMSE* and *rMAE*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

5.1.6 Bayesian Ridge regression (*BR*)

This estimator class consists of Bayesian Ridge regression estimators. All estimators from this class consist of 24 sub-estimators, each predicting a single hour.

5.1.6.1 Optimization

The parameters α_k, α_θ, λ_k, and λ_θ were optimized for each of the estimator using the NM method, the starting point for optimization was found using the PSO for model `BR_T2,7,14,28,M7,30_w-WMY`. The PSO used 20 particles for 25 generations and the parameter search space was $\alpha_k = 10^{a_k}$, $a_k \in [-8, -2]$, $\alpha_\theta = 10^{a_\theta}$, $a_\theta \in [-8, -2]$, $\lambda_k = 10^{l_k}$, $l_k \in [-8, -2]$, and $\lambda_\theta = 10^{l_\theta}$, $l_\theta \in [-8, -2]$. The initial point for the NM optimization was $a_k \approx -5.241$, $a_\theta \approx -4.295$, $l_k \approx -2.002$, and $l_\theta \approx -5.618$.

5.1.6.2 Best estimators

The 10 best estimators were selected by both MAE and RMSE as shown in table C.10 and table C.9 for further analysis using the DM test. The best estimators by Pareto rank are shown in table 5.5 which also contains most of the estimators selected by both measures.

The results of the DM test at 5% significance level are shown in fig. C.1e. When comparing the estimators selected by RMSE, there are significant differences only for group of the first three estimators compared to last four estimator. The few best estimator in the MAE selection are significanlty better than the others as shown in section C.1.

While the inclusion of weather data had a almost no effect on average for RMSE while having small positive average effect for MAE as shown infigs. A.1, A.2, D.1 and D.2 — only one estimator in the top 10 by RMSE used the weather data, but four in the top 10 by MAE and two of them were at the first positions. This estimator is quite robus against overfitting as virtually no overfitting occured as shown in fig. B.1 and fig. B.2.

Overall, the `BR` is quite good as it is on 4th position when comparing best estimators' forecasts selected by both RMSE and MAE as shown in figs. 5.1 and 5.2, furthermore, its best estimator by RMSE is 54th in overall RMSE rankings and the best by MAE is 61st in overall MAE rankings.

The example prediction on test data is shown in fig. E.5b (best by RMSE) and fig. E.5a (best by MAE) for 14th – 28th May 2016.

rPareto	**Var.**	**W.**	**RMSE**	**MAE**	**rRMSE**	**rMAE**
1/11	T2,7-DMY	Yes	7.3999	5.3492	1/54	1/61
2/16	T2,7-DM	Yes	7.4109	5.3806	3/77	2/90
2/15	T2,7,M7-DM	No	7.4006	5.4045	2/57	5/131
3/18	T2,M7,30-D	No	7.4110	5.4356	4/79	15/187
3/24	T2,7,M7-D	Yes	7.4266	5.3933	6/102	4/119
3/25	T2,7,M7-DY	Yes	7.4488	5.3874	13/152	3/108
4/29	T2,7-DY	Yes	7.4657	5.4103	17/180	6/143
4/24	T2,7,M7-D	No	7.4250	5.4445	5/100	16/197
4/30	T2,7,M7-DY	No	7.4381	5.4259	7/125	12/171
4/27	T2,7,M7-DMY	No	7.4400	5.4114	8/130	7/145
5/28	T2,M7,30-D	Yes	7.4427	5.4173	9/135	10/155
5/29	T2,7,M7-DM	Yes	7.4569	5.4127	16/167	9/148
5/30	T2,7,M7-DMY	Yes	7.4847	5.4118	20/202	8/146

Table 5.5: The list of best BR estimators with Pareto rank at most 5. Column *rPareto* contain the Pareto rank within the class and also the overal Pareto rank (equivalety for *rRMSE* and *rMAE*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

5.1.7 Elastic Net regression (*EN*)

This estimator class consists of Elastic Net regression estimators. All estimators from this class consist of 24 sub-estimators, each predicting a single hour.

5.1.7.1 Optimization

The parameters α and L1 ratio ρ were optimized for each of the estimator using the NM method, the starting point for optimization was found using the PSO for model `EN_T2-31_w-DMY`. The PSO used 100 particles for 50 generations and the parameter search space was $\alpha \in [0, 2]$ and $\rho \in [0, 1]$. The initial point for the NM optimization was $\alpha \approx 0.0089$ and $\rho \approx 0.8273$.

5.1.7.2 Best estimators

The 10 best estimators were selected by both MAE and RMSE as shown in table C.12 and table C.11 for further analysis using the DM test. The best estimators by Pareto rank are shown in table 5.6 which also contains most of the estimators selected by both measures.

The results of the DM test at 5% significance level are shown in fig. C.1f. The differences were significant only for the very best estimators from the top 10 by both MAE and RMSE when compared with several estimators that with lower rank. The rest of estimators does not show significant differences.

This class of estimators is one of those that benefited most from the inclusion of weather data. The average gain of inclusion of weather data is very small when using the RMSE measure in figs. A.1 and D.1 while the average gain is higher for the MAE measure in figs. A.2 and D.2. However, the top performing estimators from this class benefited almost always from the inclusion of the weather data as seven estimators from the top 10 by RMSE and all estimatros from the top 10 by MAE use weather data.

This estimator is quite robus against overfitting as virtually no overfitting occured as shown in fig. B.1 and fig. B.2.

Overall, the `EN` is very good as it is on 2nd position when comparing best estimators' forecasts selected by MAE and 3rd when comparing by RMSE as shown in figs. 5.1 and 5.2, furthermore, its best estimator by RMSE is 7th in overall RMSE rankings and the best by MAE is 6th in overall MAE rankings. The only estimators that were better was the `Lasso` estimator and its exact variant `LassoLars`, however, the `Lasso` estimator is a special case of parametrization of `EN` when $\rho = 1$. It is possible that the `EN` could outperform the `Lasso` when parameter optimization was done using the best performing variable combination `T2,M7,30_w-DMY` (RMSE) or `T2,7,M7_w-DMY` (MAE), however, it would be computationally very expensive to run the PSO optimization for many (or all) variable combinations.

The example prediction on test data is shown in fig. E.6b (best by RMSE) and fig. E.6a (best by MAE) for 14th – 28th May 2016.

5.1.8 Kernel Ridge Regression with linear kernel (*KRR-linear*)

This estimator class consists of Kernel Ridge regression estimators with linear kernel. All estimators from this class consist of 24 sub-estimators, each predicting a single hour. The parameter α wa optimized for each of the estimator using the NM method, the starting point for optimization was found using the PSO for model `KRR-linear_T2,7,14,28,M7,30_w-WMY`. The PSO used 20 particles for 25 generations and

rPareto	**Var.**	**W.**	**RMSE**	**MAE**	**rRMSE**	**rMAE**
1/4	T2,7,M7-DMY	Yes	7.3662	5.2980	2/11	2/12
1/3	T2,M7,30-DMY	Yes	7.3877	5.2898	10/39	1/6
1/4	T2,7,M7-DM	Yes	7.3578	5.3083	1/7	6/22
2/6	T2,7,14,28,M7,M30-DMY	Yes	7.3855	5.3021	9/38	3/15
2/6	T2,3,7-DMY	Yes	7.3794	5.3071	4/23	5/20
2/6	T2,7-DMY	Yes	7.3728	5.3382	3/14	13/46
2/5	T2,M7,30-DM	Yes	7.3839	5.3051	7/34	4/17
3/8	T2,7,14,28,M7,M30-DM	Yes	7.3854	5.3173	8/37	8/29
3/8	T2,7,M7-DMY	No	7.3805	5.3473	6/29	17/58
3/9	T2,M7,30-DY	Yes	7.4059	5.3170	16/66	7/28
3/8	T2,7,M7-DM	No	7.3796	5.3568	5/24	19/69
4/11	T2,7,M7-DY	Yes	7.4007	5.3337	14/58	11/37
4/10	T2,7,M7-D	Yes	7.3979	5.3443	12/51	15/54
4/12	T2,M7,30-DM	No	7.3977	5.3665	11/49	23/77
4/10	T2,M7,30-D	Yes	7.3995	5.3343	13/53	12/38
4/11	T2,7,14,28,M7,M30-DY	Yes	7.4103	5.3246	20/75	9/34
5/14	T2,7,14,28,M7,M30-D	Yes	7.4073	5.3391	17/70	14/47
5/13	T2,7-DM	Yes	7.4022	5.3512	15/62	18/64
5/12	T2-7-DMY	Yes	7.4399	5.3265	34/129	10/35

Table 5.6: The list of best EN estimators with Pareto rank at most 5. Column *rPareto* contain the Pareto rank within the class and also the overal Pareto rank (equivalety for *rRMSE* and *rMAE*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

the parameter search space was $\alpha = 10^a$, $a \in [-6, 6]$. The initial point for the NM optimization was $a \approx 1.622$.

5.1.8.1 Best estimators

The 10 best estimators were selected by both MAE and RMSE as shown in table C.14 and table C.13 for further analysis using the DM test. The best estimators by Pareto rank are shown in table 5.7 which also contains most of the estimators selected by both measures.

The results of the DM test at 5% significance level are shown in fig. C.1g. The differences between the top 10 estimators were mostly significant for selection by MAE and less frequently for selection by RMSE. There were no significant differences between the top 3 by MAE and top 4 by RMSE.

This class of estimators is benefit from the inclusion of weather data in general, however, the weather data had not much influence on the performance of the top estimators by MAE and slightly negative effect on the performance of top estimators by RMSE. The average gain of inclusion of weather data is very small when using the RMSE measure in figs. A.1 and D.1 while the average gain is higher for the MAE measure in figs. A.2 and D.2. However, the top performing estimators from this class benefited almost always from the inclusion of the weather data as seven estimators from the top 10 by RMSE and all estimatros from the top 10 by MAE use weather data.

This estimator is quite robus against overfitting as virtually no overfitting occured as shown in fig. B.1 and fig. B.2.

Overall, the `KRR-linear` is mediocre withwith 10th position when comparing best estimators' forecasts selected by both MAE an RMSE as shown in figs. 5.1 and 5.2, furthermore, its best estimator by RMSE is 275th in overall RMSE rankings and the best by MAE is 243rd in overall MAE rankings.

The example prediction on test data is shown in fig. E.7b (best by RMSE) and fig. E.7a (best by MAE) for 14th – 28th May 2016.

rPareto	Var.	W.	RMSE	MAE	rRMSE	rMAE
1/46	T2,M7,30-DY	No	7.5334	5.4693	1/275	2/251
1/44	T2,M7,30-DY	Yes	7.5741	5.4644	9/364	1/243
2/47	T2,M7,30-DM	No	7.5435	5.4862	2/294	5/286
2/49	T2,M7,30-DMY	No	7.5692	5.4750	6/352	3/263
3/53	T2,M7,30-DMY	Yes	7.6152	5.4821	19/507	4/274
3/52	T2,7,14,28,M7,M30-DY	Yes	7.5758	5.4971	11/370	6/316
3/55	T2,7,14,28,M7,M30-DY	No	7.5512	5.5340	4/313	17/442
3/55	T2,7,14,28,M7,M30-DM	No	7.5441	5.5494	3/296	21/518
3/56	T2,7,14,28,M7,M30-D	Yes	7.5718	5.5160	7/358	9/362
4/58	T2,7,14,28,M7,M30-DM	Yes	7.5751	5.5240	10/366	12/399
4/57	T2-7-DMY	Yes	7.6300	5.4986	26/556	7/323
4/57	T2,7,14,28,M7,M30-DMY	No	7.5652	5.5363	5/342	18/450
4/56	T2,M7,30-DM	Yes	7.6073	5.5003	18/476	8/329
4/61	T2,7,14,28,M7,M30-DMY	Yes	7.5983	5.5215	17/440	10/390
5/60	T2,7,M7-DMY	No	7.5848	5.5282	12/394	14/421
5/63	T2,M7,30-D	Yes	7.6193	5.5228	23/526	11/396
5/60	T2,M7-DMY	No	7.5848	5.5282	12/394	14/421
5/60	T2,7,14,28,M7,M30-D	No	7.5740	5.5786	8/363	31/645

Table 5.7: The list of best KRR-linear estimators with Pareto rank at most 5. Column *rPareto* contain the Pareto rank within the class and also the overal Pareto rank (equivalety for *rRMSE* and *rMAE*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

5.1.9 Kernel Ridge Regression with poly-2 kernel (*KRR-poly-2*)

This estimator class consists of Kernel Ridge regression estimators with polynomial kernel with degree 2. All estimators from this class consist of 24 sub-estimators, each predicting a single hour.

5.1.9.1 Optimization

The parameter α was optimized for each of the estimator using the NM method, the starting point for optimization was found using the PSO for model `KRR-poly-2_T2,7,14,28,M7,30_w-WMY`. The PSO used 20 particles for 25 generations and the parameter search space was $\alpha = 10^a$, $a \in [-6, 6]$. The initial point for the NM optimization was $a \approx 1.622$.

5.1.9.2 Best estimators

The 10 best estimators were selected by both MAE and RMSE as shown in table C.16 and table C.15 for further analysis using the DM test. The best estimators by Pareto rank are shown in table 5.8 which also contains most of the estimators selected by both measures.

The results of the DM test at 5% significance level are shown in fig. C.1h. The results of the DM for the top 10 by MAE shows that significant differences were mostly only between 1st estimator and 9th estimator. The two best by RMSE are significantly different from the rest of the top 10 selection (with the exception of 7th and 8th estimators), while almost no significant differences were observed in between the 7th–10th estimators.

The `KRR-poly-2` is one of the estimators for which the inclusion of weather data had negative effect as shown in figs. A.1, A.2, D.1 and D.2. While the negative effect was not large, it was significant for most of the variable combinations as shown in figs. 5.3 and 5.4. Furthermore, none of the top 10 estimators by either measure used the weather data.

This estimator is overfitted the data slightly as shown in fig. B.1 and fig. B.2.

Overall, the `KRR-poly-2` is very similar to `KRR-linear` with 7th position when comparing best estimators' forecasts selected by RMSE and 11th by MAE as shown in figs. 5.1 and 5.2, furthermore, its best estimator by RMSE is 123rd in overall RMSE rankings and the best by MAE is 204th in overall MAE rankings.

The example prediction on test data is shown in fig. E.8b (best by RMSE) and fig. E.8a (best by MAE) for 14th – 28th May 2016.

rPareto	Var.	W.	RMSE	MAE	rRMSE	rMAE
1/30	T2,7-DM	No	7.4358	5.4503	1/123	1/204
2/39	T2,7-D	No	7.4564	5.4760	2/165	9/264
2/38	T2,7-DMY	No	7.4741	5.4550	5/193	2/216
3/41	T2,14,28-DM	No	7.4696	5.5265	3/185	23/415
3/42	T2,14,28-DMY	No	7.4720	5.4965	4/191	12/315
3/41	T2,7,M7-DM	No	7.4935	5.4613	6/214	3/236
4/42	T2,7-DY	No	7.4951	5.4673	7/218	7/249
4/44	T2,7-WMY	No	7.5388	5.4645	11/282	4/245
4/44	T2,7,M7-DMY	No	7.5309	5.4658	9/267	5/247
5/45	T2,7-WM	No	7.5323	5.4672	10/271	6/248
5/43	T2,7,M7-D	No	7.5043	5.4761	8/230	10/265

Table 5.8: The list of best KRR-poly-2 estimators with Pareto rank at most 5. Column *rPareto* contain the Pareto rank within the class and also the overal Pareto rank (equivalety for *rRMSE* and *rMAE*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

5.1.10 Least Angle Regression (*Lars*)

This estimator class consists of Least Angle Regression estimators. All estimators from this class consist of 24 sub-estimators, each predicting a single hour.

5.1.10.1 Optimization

The only parameter is the number of nonzero coeffient that should be used for the regression. This number was optimized for each of the estimators by exhaustive search over the interval $[1, 30]$. The optimal value never reached the upper limit of the interval and was usually around 10 nonzero coefficients.

5.1.10.2 Best estimators

The 10 best estimators were selected by both MAE and RMSE as shown in table C.18 and table C.17 for further analysis using the DM test. The best estimators by Pareto rank are shown in table 5.9 which also contains most of the estimators selected by both measures.

The results of the DM test at 5% significance level are shown in fig. C.1i. The results of the DM test are very similar for both selection by RMSE and MAE where the estimators are grouped to pairs (estimators with and without the weather data performs the same), each pair is significantly better than pairs below.

The inclusion of weather data has no influence at all on the performance of the `Lars` estimator as shown in figs. A.1, A.2, D.1 and D.2. The procedure of selecting the non-zero coefficients selected the variables that brings the most information which were usually the dummies or the data from day $t - 2$.

This estimator is very robust to overfitting as shown in figs. B.1 and B.2 due to the variable selection process.

Overall, the `Lars` is one of the worst estimators used in the comparison with 19th position for both MAE and RMSE as shown in figs. 5.1 and 5.2, furthermore, its best estimator by RMSE is 3142nd in overall RMSE rankings and the best by MAE is 3070th in overall MAE rankings.

The example prediction on test data is shown in fig. E.9b (best by RMSE) and fig. E.9a (best by MAE) for 14th – 28th May 2016.

5.1.11 Least absolute shrinkage and selection operator (*Lasso*)

This estimator class consists of Lasso regression estimators. All estimators from this class consist of 24 sub-estimators, each predicting a single hour.

5.1.11.1 Optimization

The parameter α was optimized for each of the estimator using the NM method, the starting point for optimization was found using the PSO for model `Lasso_T2,7,14 ,28,M7,30_w-WMY`. The PSO used 20 particles for 50 generations and the parameter

rPareto	Var.	W.	RMSE	MAE	rRMSE	rMAE
1/321	nl-DMY	No	9.1701	6.7607	1/3142	1/3070
1/321	nl-DMY	Yes	9.1701	6.7607	1/3142	1/3070
2/337	nl-WMY	No	9.3553	6.9675	3/3232	3/3232
2/337	nl-WMY	Yes	9.3553	6.9675	3/3232	3/3232
3/346	nl-MY	No	9.4699	7.0923	5/3289	5/3360
3/346	nl-MY	Yes	9.4699	7.0923	5/3289	5/3360
4/368	T2-MY	No	9.7349	7.3894	7/3439	7/3574
4/368	T2,7-MY	No	9.7349	7.3894	7/3439	7/3574
5/383	T2-DMY	No	9.8209	7.4871	9/3533	9/3692
5/383	T2-DMY	Yes	9.8209	7.4871	9/3533	9/3692
5/383	T2,7-DMY	No	9.8209	7.4871	9/3533	9/3692
5/383	T2,7-DMY	Yes	9.8209	7.4871	9/3533	9/3692

Table 5.9: The list of best Lars estimators with Pareto rank at most 5. Column *rPareto* contain the Pareto rank within the class and also the overal Pareto rank (equivalety for *rRMSE* and *rMAE*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

search space was $\alpha = 10^a$, $a \in [-6, 6]$. The initial point for the NM optimization was $a \approx -2.411$.

5.1.11.2 Best estimators

The 10 best estimators were selected by both MAE and RMSE as shown in table C.20 and table C.19 for further analysis using the DM test. The best estimators by Pareto rank are shown in table 5.10 which also contains most of the estimators selected by both measures.

The results of the DM test at 5% significance level are shown in fig. C.1j. The results of DM tests shows that the differences between the estimators are mostly not significant, the major exception is the 1st estimator, which is significantly better than four (MAE) and six (RMSE) other estimators. The inclusion of weather data was beneficial for `Lasso` estimators as shown in figs. A.1, A.2, D.1 and D.2. While the effect on average was positive but very small (but still mostly significant) for the RMSE measure (fig. A.1), all of the top 10 by MAE and 8 of the top 10 by RMSE estimators used the weather data (tables C.19 and C.20). The inclusion of weather data has generally always positive effect on the MAE measurure and it was significant for about 90% of `Lasso` estimators. It is interesting that the `Lasso` version with exact solution `LassoLars` had not benefited as clearly from the inclusion of weather data as the numerical optimization `Lasso` (but still it benefited in most cases).

This estimator is quite robust to overfitting as shown in figs. B.1 and B.2 due to the variable selection process.

Overall, the `Lasso` is very good as it is on 3rd position when comparing best estimators' forecasts selected by MAE (fig. 5.1) and 2nd when comparing by RMSE (fig. 5.2), furthermore, its best estimator by RMSE is 6th in overall RMSE rankings and the best by MAE is 8th in overall MAE rankings. The version `LassoLars` with exact solution was better in the overall rankings for both RMSE and MAE which suggests that the numerical optimization using coordinate descent fail to reach global optima but finds solution quite close (but yielding significantly different forecasts). This might be caused either be existence of close local optima with objective function value close to the optimal solution or too slow convergence to local optima (and maximum iteration limit is reached before convergence, used models used limit of 1000 iterations). Another close contestant was the Elastic Net `EN` which is a generalization of `Lasso`estimators — `EN` with parameter $\rho = 1$ is equivalent to `Lasso`. The example prediction on test data is shown in fig. E.10b (best by RMSE) and fig. E.10a (best by MAE) for 14th – 28th May 2016.

rPareto	Var.	W.	RMSE	MAE	rRMSE	rMAE
1/3	T2,7,M7-DMY	Yes	7.3647	5.2977	2/10	2/10
1/3	T2,7,M7-DM	Yes	7.3566	5.3081	1/6	6/21
1/4	T2,M7,30-DMY	Yes	7.3896	5.2937	10/42	1/8
2/5	T2,7,14,28,M7,M30-DMY	Yes	7.3848	5.3014	8/35	3/14
2/5	T2,3,7-DMY	Yes	7.3785	5.3057	4/21	5/19
2/5	T2,M7,30-DM	Yes	7.3838	5.3052	7/33	4/18
2/7	T2,7,M7-DM	No	7.3758	5.3554	3/17	21/68
3/7	T2,7,14,28,M7,M30-DM	Yes	7.3848	5.3166	9/36	8/26
3/9	T2,7,M7-DMY	No	7.3816	5.3490	6/30	18/60
3/8	T2-7-DMY	Yes	7.4069	5.3113	19/69	7/23
3/7	T2,7-DMY	Yes	7.3791	5.3510	5/22	19/62
4/8	T2,M7,30-DY	Yes	7.4050	5.3166	17/65	9/27
4/10	T2,M7,30-D	Yes	7.3999	5.3332	13/55	11/36
4/10	T2,7,M7-D	Yes	7.3945	5.3445	11/46	16/55
5/12	T2,7,M7-DY	Yes	7.4016	5.3353	15/60	13/40
5/11	T2,7-DM	Yes	7.3963	5.3511	12/48	20/63
5/10	T2,7,14,28,M7,M30-DY	Yes	7.4091	5.3234	21/74	10/33

Table 5.10: The list of best Lasso estimators with Pareto rank at most 5. Column *rPareto* contain the Pareto rank within the class and also the overal Pareto rank (equivalety for *rRMSE* and *rMAE*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

5.1.12 Least absolute shrinkage and selection operator optimized using LARS (*LassoLars*)

This estimator class consists of Lasso regression estimators optimized using modified LARS (viz section 3.2.6.1). All estimators from this class consist of 24 subestimators, each predicting a single hour.

5.1.12.1 Optimization

The parameter α was optimized for each of the estimator using the NM method, the starting point for optimization was found using the PSO for model `LassoLars_T2,7,14,28,M7,30_w-WMY`. The PSO used 20 particles for 50 generations and the parameter search space was $\alpha = 10^a$, $a \in [-6, 6]$. The initial point for the NM optimization was $a \approx -3.8499$.

5.1.12.2 Best estimators

The 10 best estimators were selected by both MAE and RMSE as shown in table C.22 and table C.21 for further analysis using the DM test. The best estimators by Pareto rank are shown in table 5.11 which also contains most of the estimators selected by both measures.

The results of the DM test at 5% significance level are shown in fig. C.1k. The results of DM tests are slightly different for the RMSE and MAE measure. While the DM test (power $p = 1$) for the top 10 by MAE shows clear differences between estimatorsthat are not exactly below each other in the ranking, the DM test (power $p = 2$) for the top 10 by RMSE shows only few significant differences between the very best estimator and 3rd to 10th estimators and also betwen the 10th and 1st to 4th estimators.

The inclusion of weather data was beneficial for `LassoLars` estimators as shown in figs. A.1, A.2, D.1 and D.2. While the effect on average was positive but very small (but still mostly significant) for the RMSE measure (fig. A.1), 8 of the top 10 by both RMSE and all of the top 10 by MAE estimators used the weather data (tables C.21 and C.22). Compared to its numerical optimization version `Lasso`, the inclusion of weather had negative significant effect for more estimators (but still < 10 estimators have shown negative effect).

This estimator is quite robust to overfitting as shown in figs. B.1 and B.2 due to the variable selection process.

Overall, the `LassoLars` was the best performing estimator ranking 1st in both MAE and RMSE . The estimator `LassoLars_T2,7,M7_w-DM` ranked 1st by RMSE and 11th by MAE and the estimator `LassoLars_T2,7,14,28,M7,M30_w-DMY` ranked 1st by MAE and 5th by MAE. The other estimators with Pareto Rank 1 are `LassoLars_T2,7,14,28,M7,M30_w-DMY` (3rd by both MAE and RMSE) and `LassoLars_T2,M7,30_w-DM` (2nd by RMSE and 4th by MAE).

The example prediction on test data is shown in fig. E.11b (best by RMSE) and fig. E.11a (best by MAE) for 14th – 28th May 2016.

5.1.13 Ordinary Least Squares linear regression (*OLS*)

This estimator class consists of linear regression estimators estimated using OLS. All estimators from this class consist of 24 sub-estimators, each predicting a single hour.

5.1.13.1 Optimization

The estimator has no parameters that would have to be optimized using validation data.

5.1.13.2 Best estimators

The 10 best estimators were selected by both MAE and RMSE as shown in table C.24 and table C.23 for further analysis using the DM test. The best estimators by Pareto

rPareto	Var.	W.	RMSE	MAE	rRMSE	rMAE
1/1	T2,7,14,28,M7,M30-DMY	Yes	7.3558	5.2710	5/5	1/1
1/1	T2,7,14,28,M7,M30-DM	Yes	7.3491	5.2814	3/3	3/3
1/1	T2,7,M7-DM	Yes	7.3363	5.2980	1/1	8/11
1/1	T2,M7,30-DM	Yes	7.3465	5.2829	2/2	4/4
2/2	T2,7,M7-DMY	Yes	7.3533	5.2887	4/4	5/5
2/2	T2,M7,30-DMY	Yes	7.3689	5.2763	9/13	2/2
3/3	T2-7-DMY	Yes	7.3798	5.2931	16/26	6/7
3/5	T2,7-DM	Yes	7.3613	5.3367	7/9	18/43
3/3	T2,3,7-DMY	Yes	7.3687	5.2946	8/12	7/9
3/5	T2,7,M7-DM	No	7.3596	5.3405	6/8	21/49
4/5	T2,7,14,28,M7,M30-DY	Yes	7.3896	5.3012	23/43	9/13
4/5	T2,7,M7-DMY	No	7.3746	5.3356	10/15	17/42
4/5	T2,7,14,28,M7,M30-D	Yes	7.3783	5.3136	14/20	11/24
4/5	T2,3,7-DM	Yes	7.3768	5.3161	12/18	12/25
5/6	T2,7,M7-D	Yes	7.3746	5.3371	11/16	19/44
5/6	T2,7,M7-DY	Yes	7.3780	5.3203	13/19	13/30
5/7	T2,M7,30-DY	Yes	7.3923	5.3023	25/45	10/16

Table 5.11: The list of best LassoLars estimators with Pareto rank at most 5. Column *rPareto* contain the Pareto rank within the class and also the overal Pareto rank (equivalety for *rRMSE* and *rMAE*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rank are shown in table 5.12 which also contains most of the estimators selected by both measures.

The results of the DM test at 5% significance level are shown in fig. C.11. The results of DM tests are quite different for the RMSE and MAE measure. While the DM test (power $p = 1$) for the top 10 by MAE shows clear differences only between estimators that are further away from each other in the ranking. The DM test (power $p = 2$) for the top 10 by RMSE shows only few significant differences between estimators that are even further away from each other in the ranking.

The inclusion of weather data had negative effect on the OLS estimators mostly because overfitting to the weather data. The distribution of the RMSE and MAE with dependency on the weather is shown in figs. A.1 and A.2, the pairwise comparison between identical estimators with and without the weather variables is shown in figs. D.1 and D.2 — the rise in the error after inclusion of weather variables is clearly visible for most of the estimators.

This estimator is one of the most overfitting estimators used in this comparison as shown in figs. B.1 and B.2. Unlike most of the others estimators, it employed no regularization, thus overfitting was the price for the simplicity of the model. The only estimator that overfitted even more was the `RF`.

Overall, the `OLS` is a quite good estimator with 7th position by MAE and 8th by RMSE shown in figs. 5.1 and 5.2, furthermore, its best estimator by RMSE is 133rd in overall RMSE rankings and the best by MAE is 198th in overall MAE rankings. This difference between the RMSE and MAE performance is caused by optimizing with respect to only the RMSE and employing no regularization. The

OLS is very simple estimator that is not robust — it is very sensitive to outliers. However despite its simplicity it performed very well with the right selection of variables when comparing the estimators by RMSE.

The example prediction on test data is shown in fig. E.12b (best by RMSE) and fig. E.12a (best by MAE) for 14th – 28th May 2016.

rPareto	Var.	W.	RMSE	MAE	rRMSE	rMAE
1/31	T2,7-DM	Yes	7.4408	5.4521	1/133	2/206
1/35	T2,7-DMY	Yes	7.4673	5.4445	5/182	1/198
2/39	T2,M7,30-D	No	7.4567	5.4729	4/166	3/258
2/37	T2,7,M7-D	No	7.4487	5.4886	2/151	5/291
3/38	T2,7-DM	No	7.4497	5.4905	3/155	6/299
3/40	T2,7-DMY	No	7.4689	5.4784	6/184	4/271
4/44	T2,3,7-DM	No	7.4903	5.5182	7/210	9/371
4/45	T2,3,7-DMY	No	7.5012	5.5108	10/228	7/344
5/51	T2,3,7-DY	No	7.5258	5.5380	11/259	12/460
5/51	T2,7-DY	Yes	7.5373	5.5175	17/281	8/368
5/46	T2,M7-D	No	7.4972	5.5406	8/223	14/475
5/50	T2,7,M7-D	Yes	7.5305	5.5244	12/266	10/402

Table 5.12: The list of best OLS estimators with Pareto rank at most 5. Column *rPareto* contain the Pareto rank within the class and also the overal Pareto rank (equivalety for *rRMSE* and *rMAE*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

5.1.14 Random Forest regression (*RF*)

This estimator class consists of Random Forest estimators each with 2000 weak estimators per single sub-estimator. All estimators from this class consist of 24 sub-estimators, each predicting a single hour.

5.1.14.1 Optimization

The individual parameters were not optimized out of the box but rather based on preliminary experiments with different parametrizations — the number of estimators is llimited only by compuational power. The maximum depth was optimized using grid search for estimator `RF_T2-14,M7,30_p-DM`.

5.1.14.2 Best estimators

The 10 best estimators were selected by both MAE and RMSE as shown in table C.26 and table C.25 for further analysis using the DM test. The best estimators by Pareto rank are shown in table 5.13 which also contains most of the estimators selected by both measures.

The results of the DM test at 5% significance level are shown in fig. C.1m. The results of DM tests are quite different for the RMSE and MAE measure. The DM test (power $p = 1$) for the top 10 by MAE shows clear differences only for the 1st and 2nd compared t some of the other estimators. The DM test (power $p = 2$) for the top 10 by RMSE shows only few significant differences between the first half and the second half with no significant differences between the `RF_T2_p-DY` and `RF_T2,7_p-DY` compared to the other estimators.

The inclusion of weather data had mostly negative effect on the RF estimators mostly because overfitting to the weather data (similarly to OLS). The distribution of the RMSE and MAE with dependency on the weather is shown in figs. A.1 and A.2, the pairwise comparison between identical estimators with and without the weather variables is shown in figs. D.1 and D.2 — compared to most of the used estimators, it shows quite a spread as the weather data had large effect (both positive and negative) on several estimators. On the best performing estimators the effect was very small but still negative.

This estimators was the most overfitting estimator used in the comparison figs. B.1 and B.2. Despite the overfitting, the estimators was still able to perform comparatively well even though its performance is in the lower half of the comparison.

Overall, the `RF` is in the lower half of estimators used in the comparison with 15th position by both RMSE and MAE as shown in figs. 5.1 and 5.2, furthermore, its best estimator by RMSE is 2425th in overall RMSE rankings and the best by MAE is 2549th in overall MAE rankings.

The example prediction on test data is shown in fig. E.13b (best by RMSE) and fig. E.13a (best by MAE) for 14th – 28th May 2016.

rPareto	**Var.**	**W.**	**RMSE**	**MAE**	**rRMSE**	**rMAE**
1/228	T2-DY	No	8.2471	6.1772	1/2425	36/2549
1/228	T2,7-DY	No	8.2471	6.1772	1/2425	36/2549
1/225	T2,14,28-D	No	8.2574	6.0914	3/2442	1/2391
2/229	T2-7-W	No	8.4305	6.0953	48/2617	2/2400
2/226	T2,14,28-DY	No	8.2641	6.0984	4/2450	3/2409
3/231	T2-7-WMY	No	8.4214	6.1010	46/2612	4/2420
3/228	T2,14,28-W	No	8.2651	6.1028	5/2454	5/2427
4/232	T2-7-WY	No	8.4355	6.1033	53/2626	6/2428
4/229	T2,14,28-WM	No	8.2727	6.1084	7/2462	10/2442
4/229	T2,14,28-DM	No	8.2718	6.1146	6/2460	13/2449
4/229	T2,14,28-DMY	No	8.2854	6.1039	10/2479	7/2429
5/232	T2-7-D	No	8.4334	6.1055	52/2624	8/2433
5/230	T2,14,28-WMY	No	8.2794	6.1103	8/2470	12/2444

Table 5.13: The list of best RF estimators with Pareto rank at most 5. Column *rPareto* contain the Pareto rank within the class and also the overal Pareto rank (equivalety for *rRMSE* and *rMAE*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

5.1.15 Support Vector Regression with linear kernel and ϵ-insensitive loss (*SVR-linear-e*)

This estimator class consists of Support Vector Regresssion with linear kernel and ϵ-insensitive loss. All estimators from this class consist of 24 sub-estimators, each predicting a single hour.

5.1.15.1 Optimization

The parameters ϵ and C were optimized for each of the estimator using the NM method, the starting point for optimization was found using the PSO for model `SVR-linear-e_T2,7,14,28,M7,30_w-WMY`. The PSO used 100 particles for 30 generations and the parameter search space was $\epsilon \in [0, 1]$ and $C = 10^c$, $c \in [-6, 2]$. The initial point for the NM optimization was $\epsilon \approx 0.3627$ and $c \approx -1.7564$.

5.1.15.2 Best estimators

The 10 best estimators were selected by both MAE and RMSE as shown in table C.28 and table C.27 for further analysis using the DM test. The best estimators by Pareto rank are shown in table 5.14 which also contains most of the estimators selected by both measures.

The results of the DM test at 5% significance level are shown in fig. C.1n. The results of DM tests are quite different for the RMSE and MAE measure. The DM test (power $p = 1$) for the top 10 by MAE shows significant differences between the estimators. The DM test (power $p = 2$) for the top 10 by RMSE shows significant differences mainly for the 1st estimator compared to the other.

The inclusion of weather data had mostly positive effect on the SVR-linear-e estimators as shown in figs. A.1, A.2, D.1 and D.2 even though the effect was quite small and not present for all estimators — especially for the top estimators. The top 10 by RMSE contains only one estimator using the weather data and the top 10 by MAE contained 5 estimators using the weather data.

This estimators was quite robust to overfitting figs. B.1 and B.2 even though not as robust as Lasso, `LassoLars`, `BR`, and `EN`.

Overall, the `SVR-linear-e` is very good estimator with 5th position by both MAE and RMSE as shown in figs. 5.1 and 5.2, furthermore, its best estimator by RMSE is 136th in overall RMSE rankings and the best by MAE is 120th in overall MAE rankings. It is the best performing SVR used in the comparison closely followed by `SVR-linear-e2` which only employs squared ϵ-insensitive loss.

The example prediction on test data is shown in fig. E.14b (best by RMSE) and fig. E.14a (best by MAE) for 14th – 28th May 2016.

rPareto	**Var.**	**W.**	**RMSE**	**MAE**	**rRMSE**	**rMAE**
1/25	T2,M7,30-D	No	7.4435	5.3998	1/136	2/127
1/28	T2,M7,30-DY	No	7.5117	5.3953	5/242	1/120
2/32	T2,M7,30-D	Yes	7.5238	5.4158	7/257	4/153
2/34	T2,M7,30-DM	No	7.4890	5.4308	2/208	5/177
2/30	T2,M7,30-DY	Yes	7.5472	5.4045	12/308	3/132
3/46	T2,M7,30-DMY	No	7.5339	5.4691	10/276	8/250
3/44	T2,M7,30-DMY	Yes	7.5919	5.4567	24/417	6/224
3/44	T2,7,M7-DM	No	7.5011	5.4831	3/227	9/277
4/48	T2-7-DMY	Yes	7.6332	5.4651	40/565	7/246
4/46	T2,M7-DM	No	7.5065	5.4987	4/233	13/324
4/50	T2,M7,30-DM	Yes	7.5772	5.4861	22/373	10/283
4/50	T2,7,14,28,M7,M30-D	Yes	7.5668	5.4920	18/344	12/301
5/54	T2-7-DM	Yes	7.6578	5.4865	48/648	11/287
5/48	T2,M7-DY	No	7.5232	5.5055	6/255	16/336
5/57	T2,7,14,28,M7,M30-DY	Yes	7.6016	5.5027	27/453	15/334

Table 5.14: The list of best SVR-linear-e estimators with Pareto rank at most 5. Column *rPareto* contain the Pareto rank within the class and also the overal Pareto rank (equivalety for *rRMSE* and *rMAE*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

5.1.16 Support Vector Regression with linear kernel and squared ϵ-insensitive loss (*SVR-linear-e2*)

This estimator class consists of Support Vector Regresssion with linear kernel and squared ϵ-insensitive loss. All estimators from this class consist of 24 sub-estimators, each predicting a single hour.

5.1.16.1 Optimization

The parameters ϵ and C were optimized for each of the estimator using the NM method, the starting point for optimization was found using the PSO for model `SVR-linear-e2_T2,7,14,28,M7,30_w-WMY`. The PSO used 100 particles for 30 generations and the parameter search space was $\epsilon \in [0, 1]$ and $C = 10^c$, $c \in [-6, 2]$. The initial point for the NM optimization was $\epsilon \approx 0.0001$ and $c \approx -1.9492$.

5.1.16.2 Best estimators

The 10 best estimators were selected by both MAE and RMSE as shown in table C.30 and table C.29 for further analysis using the DM test. The best estimators by Pareto rank are shown in table 5.15 which also contains most of the estimators selected by both measures.

The results of the DM test at 5% significance level are shown in fig. C.1o. The DM tests show only scattered significant differences — the 1st estimator is significantly better than at least half of the others estimators for both RMSE and MAE.

Similarly as for `SVR-linear-e`, the inclusion of weather data had mostly positive effect on the `SVR-linear-e2` estimators as shown in figs. A.1, A.2, D.1 and D.2 even though the effect was quite small and not present for all estimators — especially for the top estimators. The top 10 by RMSE contains only one estimator using the weather data and the top 10 by MAE contained 5 estimators using the weather data.

This estimators was quite robust to overfitting figs. B.1 and B.2 even though not as robust as Lasso, `LassoLars`, `BR`, and `EN`.

Overall, the `SVR-linear-e2` is quite good estimator with 9th position by RMSE and 6th by MAE as shown in figs. 5.1 and 5.2, furthermore, its best estimator by RMSE is 186th in overall RMSE rankings and the best by MAE is 193rd in overall MAE rankings. It is the second best performing SVR estimator but worse than the version with ϵ-insensitve loss `SVR-linear-e`.

The example prediction on test data is shown in fig. E.15b (best by RMSE) and fig. E.15a (best by MAE) for 14th – 28th May 2016.

rPareto	**Var.**	**W.**	**RMSE**	**MAE**	**rRMSE**	**rMAE**
1/38	T2,M7,30-D	No	7.4698	5.4589	1/186	7/232
1/34	T2,M7,30-DY	No	7.4848	5.4391	2/203	2/194
1/39	T2,M7,30-DY	Yes	7.5412	5.4376	14/288	1/193
2/39	T2,M7,30-D	Yes	7.5237	5.4476	8/256	3/200
2/36	T2,M7,30-DM	No	7.4953	5.4485	3/219	4/202
3/40	T2,M7,30-DMY	No	7.5369	5.4539	12/279	5/211
3/45	T2,7,M7-DM	No	7.5054	5.4890	4/231	10/295
4/47	T2,M7-DM	No	7.5151	5.4946	5/248	12/307
4/42	T2,M7,30-DM	Yes	7.5549	5.4614	19/322	8/237
4/45	T2,M7,30-DMY	Yes	7.5814	5.4583	29/384	6/228
5/53	T2-7-DMY	Yes	7.6146	5.4862	37/506	9/285
5/48	T2,M7-DMY	No	7.5414	5.4984	15/289	13/321
5/48	T2,7,M7-DY	No	7.5187	5.5110	6/251	16/345
5/51	T2,7,14,28,M7,M30-DY	Yes	7.5685	5.4920	23/347	11/302

Table 5.15: The list of best SVR-linear-e2 estimators with Pareto rank at most 5. Column *rPareto* contain the Pareto rank within the class and also the overal Pareto rank (equivalety for *rRMSE* and *rMAE*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

5.1.17 Support Vector Regression with polynomial kernel of degree 2 (*SVR-poly-2*)

This estimator class consists of Support Vector Regresssion with polynomial kernel of degree 2 . All estimators from this class consist of 24 sub-estimators, each predicting a single hour.

5.1.17.1 Optimization

The parameters ϵ and C were optimized for each of the estimator using the NM method, the starting point for optimization was found using the PSO for model `SVR-poly-2_T2,7,14,28,M7,30_w-WMY`. The PSO used 100 particles for 30 generations and the parameter search space was $\epsilon \in [0, 1]$ and $C = 10^c$, $c \in [-6, 2]$. The initial point for the NM optimization was $\epsilon \approx 0.0760$ and $c \approx 0.6027$.

5.1.17.2 Best estimators

The 10 best estimators were selected by both MAE and RMSE as shown in table C.32 and table C.31 for further analysis using the DM test. The best estimators by Pareto rank are shown in table 5.16 which also contains most of the estimators selected by both measures.

The results of the DM test at 5% significance level are shown in fig. C.1p. The DM tests have slightly different results for both MAE and RMSE, when the first three estimators are significantly better than the others by the MAE and only the very first estimator is significantly better than the others by RMSE. The DM tests for RMSE also show significant dominance over the last four estimators by the first three estimators.

In contrast to `SVR-linear-e` and `SVR-linear-e`, the inclusion of weather data had negative effect on the `SVR-poly-2` estimators as shown in figs. A.1, A.2, D.1 and D.2 even though the effect was quite small and not present for all estimators — especially for the top estimators. The top 10 by RMSE contains only one estimator using the weather data and the top 10 by MAE contained no estimators using the weather data.

This estimators was one of the estimators that overfitted for most variable combinations as shown in figs. B.1 and B.2.

Overall, the `SVR-poly-2` is lower half of the estimators. It ranks 14th by both RMSE and MAE as shown in figs. 5.1 and 5.2, furthermore, its best estimator by RMSE is 1953rd in overall RMSE rankings and the best by MAE is 2126th in overall MAE rankings. Other SVR with polynomial kernel of degree 3 and 4 performed even worse.

The example prediction on test data is shown in fig. E.16b (best by RMSE) and fig. E.16a (best by MAE) for 14th – 28th May 2016.

5.1.18 Support Vector Regression with polynomial kernel of degree 3 (*SVR-poly-3*)

This estimator class consists of Support Vector Regresssion with polynomial kernel of degree 3 . All estimators from this class consist of 24 sub-estimators, each predicting a single hour.

rPareto	Var.	W.	RMSE	MAE	rRMSE	rMAE
1/180	T2-DM	No	8.0146	5.9466	1/1953	1/2126
2/202	T2,7-WY	No	8.1896	5.9930	3/2346	3/2241
2/197	T2,7-DY	No	8.2076	5.9673	4/2378	2/2173
2/213	T2,14,28-DM	No	8.1882	6.0786	2/2342	5/2373
3/211	T2-DY	No	8.2965	6.0383	9/2497	4/2316
3/224	T2-DMY	No	8.2641	6.0855	7/2452	7/2382
3/224	T2,7-DM	No	8.2891	6.0831	8/2487	6/2377
3/223	T2,14,28-DMY	No	8.2227	6.0973	5/2402	8/2404
4/230	T2,7-DMY	No	8.3250	6.0985	15/2527	9/2411
4/229	T2,14,28-WM	No	8.2563	6.1542	6/2441	12/2500
4/235	T2,7,14,28,M7,M30-DMY	No	8.3158	6.1429	12/2517	11/2491
5/236	T2,7-WM	No	8.3183	6.1557	13/2518	13/2503
5/236	T2,7-WM	Yes	8.3014	6.1642	10/2500	15/2522
5/236	T2,7-WMY	No	8.3816	6.1358	18/2585	10/2480

Table 5.16: The list of best SVR-poly-2 estimators with Pareto rank at most 5. Column *rPareto* contain the Pareto rank within the class and also the overal Pareto rank (equivalety for *rRMSE* and *rMAE*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

5.1.18.1 Optimization

The parameters ϵ and C were optimized for each of the estimator using the NM method, the starting point for optimization was found using the PSO for model `SVR-poly-3_T2,7,14,28,M7,30_w-WMY`. The PSO used 100 particles for 30 generations and the parameter search space was $\epsilon \in [0, 1]$ and $C = 10^c$, $c \in [-6, 3]$. The initial point for the NM optimization was $\epsilon \approx 0.6408$ and $c \approx 0.1202$.

5.1.18.2 Best estimators

The 10 best estimators were selected by both MAE and RMSE as shown in table C.34 and table C.33 for further analysis using the DM test. The best estimators by Pareto rank are shown in table 5.17 which also contains most of the estimators selected by both measures.

The results of the DM test at 5% significance level are shown in fig. C.1q. The DM test shows significant differences between most of the estimators for both MAE and RMSE selections.

Similarly as `SVR-linear-e` and `SVR-linear-e` and in contrast to `SVR-poly-2` and `SVR-poly-4` , the inclusion of weather data had positive effect on the `SVR-poly-3` estimators as shown in figs. A.1, A.2, D.1 and D.2. All estimator of the top 10 by RMSE estimator and 8 estimators of the top 10 by MAE use the weather data.

This estimators was one of the estimators that overfitted for most variable combinations as shown in figs. B.1 and B.2 — this behavior is present also for `SVR-poly-2` and `SVR-poly-4`.

Overall, the `SVR-poly-3` is in one of the worst estimators — it ranks 17th by both RMSE and MAE as shown in figs. 5.1 and 5.2, furthermore, its best estimator by RMSE is 2892nd in overall RMSE rankings and the best by MAE is 2851st in overall MAE rankings. The estimator with polynomial degree 4 performed even worse.

The example prediction on test data is shown in fig. E.17b (best by RMSE) and fig. E.17a (best by MAE) for 14th – 28th May 2016.

rPareto	Var.	W.	RMSE	MAE	rRMSE	rMAE
1/292	T2,7,M7-DY	Yes	8.8580	6.5185	1/2892	1/2851
2/295	T2,7,M7-WY	Yes	8.8978	6.5763	2/2917	2/2883
3/299	T2,7,M7-D	Yes	8.9203	6.5799	3/2940	3/2885
4/304	T2,7,M7-W	Yes	8.9586	6.6344	4/2979	6/2952
4/302	T2,7,M7-DMY	Yes	8.9654	6.6024	5/2984	4/2911
5/303	T2,7,M7-DM	Yes	8.9811	6.6288	6/2995	5/2945

Table 5.17: The list of best SVR-poly-3 estimators with Pareto rank at most 5. Column *rPareto* contain the Pareto rank within the class and also the overal Pareto rank (equivalety for *rRMSE* and *rMAE*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

5.1.19 Support Vector Regression with polynomial kernel of degree 4 (*SVR-poly-4*)

This estimator class consists of Support Vector Regresssion with polynomial kernel of degree 4. All estimators from this class consist of 24 sub-estimators, each predicting a single hour.

5.1.19.1 Optimization

The parameters ϵ and C were optimized for each of the estimator using the NM method, the starting point for optimization was found using the PSO for model `SVR-poly-4_T2,7,14,28,M7,30_w-WMY`. The PSO used 100 particles for 30 generations and the parameter search space was $\epsilon \in [0, 1]$ and $C = 10^c$, $c \in [-6, 3]$. The initial point for the NM optimization was $\epsilon \approx 0.6408$ and $c \approx 0.1202$.

5.1.19.2 Best estimators

The 10 best estimators were selected by both MAE and RMSE as shown in table C.36 and table C.35 for further analysis using the DM test. The best estimators by Pareto rank are shown in table 5.18 which also contains most of the estimators selected by both measures.

The results of the DM test at 5% significance level are shown in fig. C.1r. The DM test for the top 10 by MAE shows significant differences between the first three estimators compared to the others, on the other hand, the estimators in the top 10 by RMSE show significant differences only between the very first estimator and the rest and also between the first three and last three estimators.

The weather had almost no effect on the prediction error by both RMSE and MAE, viz figs. A.1, A.2, D.1 and D.2 — 3 estimator of the top 10 by by both MAE and RMSE use the weather data. The effect was significant only for several estimators and then the effect was mixed — positive for some and negative for others viz figs. 5.3 and 5.4.

This estimators was one of the estimators that overfitted for most variable combinations as shown in figs. B.1 and B.2 — this behavior is present also for `SVR-poly-2` and `SVR-poly-3`.

Overall, the `SVR-poly-4` is the worst estimator used in the comparison — it ranks 20th by both RMSE and MAE as shown in figs. 5.1 and 5.2, furthermore, its best estimator by RMSE is 4289th in overall RMSE rankings and the best by MAE is 3979th in overall MAE rankings.

The example prediction on test data is shown in fig. E.18b (best by RMSE) and fig. E.18a (best by MAE) for 14th – 28th May 2016.

rPareto	**Var.**	**W.**	**RMSE**	**MAE**	**rRMSE**	**rMAE**
1/437	T2,7,M7-DY	No	10.6661	7.6831	1/4289	1/3979
2/456	T2,7-DY	No	10.6754	7.8203	2/4292	3/4114
2/452	T2,7,M7-DMY	No	10.8727	7.8074	5/4395	2/4105
3/460	T2,7-DMY	No	10.7516	7.9046	3/4318	7/4164
3/459	T2,7,M7-WY	No	10.8223	7.8733	4/4382	4/4145
4/477	T2,7-Y	No	10.9365	8.1641	8/4414	24/4369
4/463	T2,7-WY	No	11.0025	7.9200	12/4447	8/4181
4/466	T2,7-DY	Yes	10.9533	7.9957	9/4418	13/4259
4/460	T2,7,M7-D	No	11.1185	7.8997	19/4543	6/4158
4/460	T2,7,M7-DMY	Yes	11.1250	7.8959	21/4546	5/4157
4/481	T2,7,14,28,M7,M30-WMY	Yes	10.8916	8.4583	6/4399	45/4517
5/475	T2,7-MY	No	10.9960	8.1044	11/4443	18/4319
5/466	T2,7-DMY	Yes	11.0202	7.9541	13/4451	10/4240
5/464	T2,7,M7-DY	Yes	11.1203	7.9338	20/4545	9/4233
5/485	T2,7,14,28,M7,M30-WMY	No	10.9900	8.4491	10/4439	44/4512
5/482	T2,7,14,28,M7,M30-DMY	Yes	10.9120	8.4669	7/4408	47/4532

Table 5.18: The list of best SVR-poly-4 estimators with Pareto rank at most 5. Column *rPareto* contain the Pareto rank within the class and also the overal Pareto rank (equivalety for *rRMSE* and *rMAE*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

5.1.20 Support Vector Regression with RBF kernel (*SVR-rbf*)

This estimator class consists of Support Vector Regresssion with RBF kernel. All estimators from this class consist of 24 sub-estimators, each predicting a single hour.

5.1.20.1 Optimization

The parameters ϵ and C were optimized for each of the estimator using the NM method, the starting point for optimization was found using the PSO for model `SVR-rbf_T2,7,14,28,M7,30_w-WMY`. The PSO used 20 particles for 50 generations and the parameter search space was $\epsilon \in [0, 1]$ and $C = 10^c$, $c \in [-6, 3]$. The initial point for the NM optimization was $\epsilon \approx 0.2230$ and $c \approx -0.0400$.

5.1.20.2 Best estimators

The 10 best estimators were selected by both MAE and RMSE as shown in table C.38 and table C.37 for further analysis using the DM test. The best estimators by Pareto rank are shown in table 5.19 which also contains most of the estimators selected by both measures.

The results of the DM test at 5% significance level are shown in fig. C.1s. The DM test shows significant differences between the first two estimators and the rest for top 10 by both RMSE and MAE.

The inclusion of weather data had negative effect as shown in figs. A.1, A.2, D.1 and D.2. Only 3 of the top 10 by RMSE estimator and only 1 estimators of the top 10 by MAE use the weather data. The effect was significant only for several estimators and then the effect was mixed — positive for some and negative for others viz figs. 5.3 and 5.4).

While this estimators also overfitted quite a lot (viz figs. B.1 and B.2), the overfitting was not as bad as for other SVR with polynomial kernel with degree 2 – 4.

Overall, the `SVR-rbf` is a mediocre estimator in the comparison — it ranks 12th by both MAE and RMSE as shown in figs. 5.1 and 5.2, furthermore, its best estimator by RMSE is 337th in overall RMSE rankings and the best by MAE is 306th in overall MAE rankings.

The example prediction on test data is shown in fig. E.19b (best by RMSE) and fig. E.19a (best by MAE) for 14th – 28th May 2016.

5.1.21 Support Vector Regression with Sigmoid kernel (*SVR-sigmoid*)

This estimator class consists of Support Vector Regresssion with Sigmoid kernel. All estimators from this class consist of 24 sub-estimators, each predicting a single hour.

5.1.21.1 Optimization

The parameters ϵ and C were optimized for each of the estimator using the NM method, the starting point for optimization was found using the PSO for model `SVR-sigmoid_T2,7,14,28,M7,30_w-WMY`. The PSO used 20 particles for 50 generations

rPareto	Var.	W.	RMSE	MAE	rRMSE	rMAE
1/57	T2-DM	Yes	7.5631	5.5590	1/337	12/559
1/55	T2-DY	No	7.6096	5.4941	2/487	3/306
1/41	T2,7-DMY	No	7.6959	5.4543	9/781	1/212
2/72	T2-DM	No	7.6371	5.5419	3/578	11/481
2/65	T2-DMY	No	7.6742	5.5242	4/703	8/401
2/52	T2,7-DM	No	7.7169	5.4783	12/855	2/270
2/60	T2,7-DY	No	7.6935	5.5126	8/770	5/351
3/69	T2,M7-DM	No	7.6809	5.5355	5/720	9/447
3/62	T2,7-WMY	No	7.7579	5.5189	18/1041	7/374
3/60	T2,7-DMY	Yes	7.7611	5.5125	19/1049	4/350
4/84	T2,14,28-DMY	No	7.6869	5.5943	7/742	21/706
4/91	T2,3,7-DM	Yes	7.6855	5.6467	6/735	50/976
4/62	T2,7-D	No	7.7732	5.5159	23/1101	6/361
5/97	T2-D	Yes	7.7061	5.6332	10/813	42/908
5/93	T2,14,28-DM	No	7.7134	5.6220	11/840	35/835
5/71	T2,7-WM	No	7.7792	5.5393	25/1123	10/469
5/97	T2-DMY	Yes	7.7547	5.6102	17/1025	29/785
5/96	T2,3,7-DM	No	7.7245	5.6213	13/890	33/831
5/95	T2,7,M7-DM	No	7.7767	5.5950	24/1116	22/709

Table 5.19: The list of best SVR-rbf estimators with Pareto rank at most 5. Column *rPareto* contain the Pareto rank within the class and also the overal Pareto rank (equivalety for *rRMSE* and *rMAE*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

and the parameter search space was $\epsilon \in [0,1]$ and $C = 10^c$, $c \in [-6,3]$. The initial point for the NM optimization was $\epsilon \approx 0.6066$ and $c \approx -1.3205$.

5.1.21.2 Best estimators

The 10 best estimators were selected by both MAE and RMSE as shown in table C.40 and table C.39 for further analysis using the DM test. The best estimators by Pareto rank are shown in table 5.20 which also contains most of the estimators selected by both measures.

The results of the DM test at 5% significance level are shown in fig. C.1t. The DM tests for top 10 by MAE show that the `T2-31-WMY` and `T2-31-DMY` are significantly better than the estimators that have the same lag variables but are missing some of the dummy variables. Furthermore, the DM tests show significant differences between most of the estimators selected by RMSE with the exception of `T2-31-WMY` and `T2-31-DMY` that are significantly dominated only by `T2,14,28-DMY` and `T2,14,28-DM`.

The inclusion of weather data had generally positive effect as shown in figs. A.1, A.2, D.1 and D.2. All estimator of the top 10 by both RMSE and MAE use the weather data.

While this estimators also overfitted quite a lot (viz figs. B.1 and B.2), the overfitting was not as bad as for other SVR with polynomial kernel with degree 2 –

4.

Overall, the `SVR-sigmoid` is one of the worst estimators in the comparison — it ranks 18th by both MAE and RMSE as shown in figs. 5.1 and 5.2, furthermore, its best estimator by RMSE is 2959th in overall RMSE rankings and the best by MAE is 2930th in overall MAE rankings.

The example prediction on test data is shown in fig. E.20b (best by RMSE) and fig. E.20a (best by MAE) for 14th – 28th May 2016.

rPareto	**Var.**	**W.**	**RMSE**	**MAE**	**rRMSE**	**rMAE**
1/307	T2-31-WMY	Yes	9.0637	6.6173	8/3042	2/2932
1/307	T2-31-DMY	Yes	9.0637	6.6168	9/3043	1/2930
1/303	T2,14,28-DMY	Yes	8.9394	6.6319	1/2959	3/2951
2/310	T2-31-M	Yes	9.0840	6.6379	11/3067	4/2953
2/305	T2,14,28-DM	Yes	8.9646	6.6427	2/2983	7/2960
3/311	T2-31-DY	Yes	9.0896	6.6399	13/3069	5/2955
3/306	T2,14,28-WMY	Yes	8.9786	6.6567	3/2994	13/2973
4/312	T2-31-WM	Yes	9.0903	6.6409	14/3071	6/2958
4/307	T2,14,28-DY	Yes	9.0105	6.6768	4/3007	19/2995
5/313	T2-31-WY	Yes	9.0974	6.6472	16/3080	8/2964
5/313	T2-31-DM	Yes	9.0921	6.6485	15/3073	10/2966
5/309	T2,14,28-MY	Yes	9.0288	6.6906	5/3021	25/3017

Table 5.20: The list of best SVR-sigmoid estimators with Pareto rank at most 5. Column *rPareto* contain the Pareto rank within the class and also the overal Pareto rank (equivalety for *rRMSE* and *rMAE*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

5.2 Comparison between classes

While the sections above shortly described the results from the point of view of individual estimators, this section contains the compares the estimator classes with each other. For purposes of this comparison, best estimator from each estimator class was selected. The best estimators by class were then compared to each other using the DM test at 5% significance level (viz figs. 5.1 and 5.2). The summary of the best results by estimator classes and the combination of variables is shown in table 5.21 (RMSE) and table 5.22 (MAE). The *LassoLars* is performing the best with almost any combination of lagged and mean variables — the only time that it is not the best is when no lagged and mean variables (named as `nl`) are present and only dummies are available to the estimators. The `ANN48-0.5d-linear` performs the best when only dummies are present for both RMSE and MAE — this suggest that the relationship between the electricity price and the dummies is not linear.

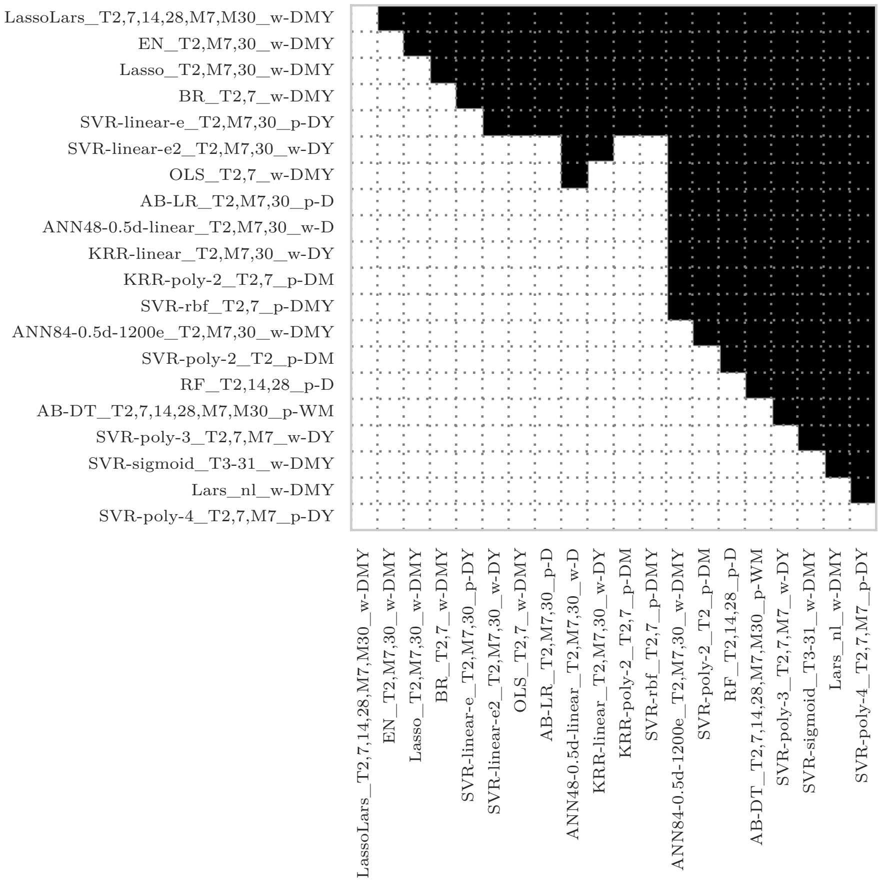


Figure 5.1: The results for DM tests (power $p = 1$) for best estimator from each class by MAE. A filled square at $[i, j]$ denotes that the forecast of estimator i is better than the forecast j by the DM test on 5% significance level.

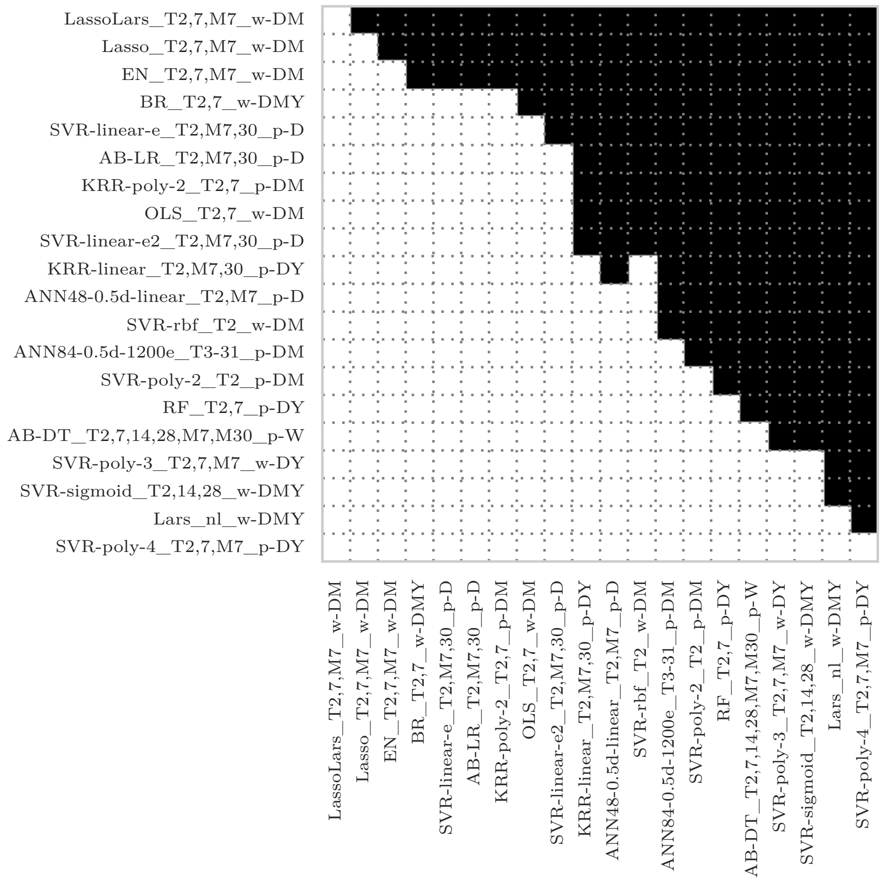


Figure 5.2: The results for DM tests (power $p = 2$) for best estimator from each class by RMSE. A filled square at $[i, j]$ denotes that the forecast of estimator i is better than the forecast j by the DM test on 5% significance level.

	T2	T2,14,28	T2,3,7	T2,7	T2,7,14,28, M7,M30	T2,7,M7	T2,M7	T2,M7,30	T2-7	T2-31	nl
AB-DT	8.945	8.838	8.935	8.948	8.620	8.632	8.642	8.637	8.887	8.774	10.740
AB-LR	7.581	7.558	7.585	7.582	7.508	7.497	7.497	7.450	7.673	9.478	9.095
ANN48-0.5d-linear	7.608	7.703	7.649	7.636	7.629	7.611	7.584	7.587	7.789	7.995	8.774
ANN84-0.5d-1200e	7.925	7.817	7.941	7.939	7.711	7.804	7.758	7.698	7.724	7.666	9.509
BR	7.577	7.595	7.612	7.400	7.515	7.401	7.456	7.411	7.585	7.686	9.095
EN	7.550	7.419	7.379	7.373	7.385	7.358	7.434	7.384	7.440	7.539	9.103
KRR-linear	7.755	7.686	7.773	7.755	7.544	7.585	7.585	7.533	7.617	7.710	10.028
KRR-poly-2	7.639	7.470	7.699	7.436	7.609	7.493	7.562	7.673	7.644	7.803	9.039
Lars	9.735	11.050	10.049	9.735	11.337	10.516	10.516	11.678	10.516	10.269	9.170
Lasso	7.552	7.418	7.379	7.379	7.385	7.357	7.432	7.384	7.407	7.572	9.515
LassoLars	7.544	7.381	7.369	7.361	7.349	**7.336**	7.401	7.347	7.380	7.512	9.515
OLS	7.585	7.500	7.490	7.441	7.537	7.449	7.497	7.457	7.686	9.465	9.098
RF	8.247	8.257	8.608	8.247	8.376	8.648	8.648	8.919	8.421	8.617	9.932
SVR-linear-e	7.597	7.569	7.642	7.600	7.560	7.501	7.507	7.443	7.633	7.709	9.099
SVR-linear-e2	7.578	7.594	7.698	7.578	7.521	7.505	7.515	7.470	7.592	7.693	9.098
SVR-poly-2	8.015	8.188	8.478	8.190	8.316	8.474	8.971	9.341	8.675	10.958	9.413
SVR-poly-3	10.390	9.459	10.296	9.072	9.080	8.858	9.405	9.426	9.397	9.265	10.702
SVR-poly-4	11.818	11.435	12.194	10.675	10.892	10.666	11.498	11.413	12.207	11.811	11.388
SVR-rbf	7.563	7.687	7.685	7.693	7.790	7.777	7.681	7.841	7.853	7.962	9.408
SVR-sigmoid	10.368	8.939	10.651	10.368	9.176	10.056	10.056	10.304	9.298	9.064	9.509

Table 5.21: The best results (RMSE) for individual estimators and variable sets. Each cell contains the best value over all possible dummies and the inclusion of weather data. The highest value for given estimator class is in green, for the particular selection of variables in blue, and the overall highest value is in **bold**.

	T2	T2,14,28	T2,3,7	T2,7	T2,7,14,28, M7,M30	T2,7,M7	T2,M7	T2,M7,30	T2-7	T2-31	nl
AB-DT	6.864	6.646	6.803	6.861	6.344	6.522	6.539	6.388	6.510	6.408	8.409
AB-LR	5.547	5.627	5.552	5.552	5.554	5.545	5.539	5.470	5.678	7.251	6.678
ANN48-0.5d-linear	5.556	5.695	5.599	5.575	5.531	5.581	5.548	5.495	5.760	5.969	6.442
ANN84-0.5d-1200e	5.830	5.732	5.821	5.863	5.600	5.741	5.688	5.560	5.566	5.689	7.089
BR	5.533	5.603	5.544	5.349	5.494	5.387	5.485	5.417	5.486	5.607	6.680
EN	5.516	5.373	5.307	5.338	5.302	5.298	5.407	5.290	5.326	5.421	6.682
KRR-linear	5.670	5.643	5.640	5.670	5.497	5.528	5.528	5.464	5.499	5.621	7.453
KRR-poly-2	5.663	5.496	5.688	5.450	5.563	5.461	5.567	5.590	5.514	5.704	6.669
Lars	7.389	8.640	7.724	7.389	8.894	8.153	8.153	9.185	8.153	7.927	6.761
Lasso	5.519	5.377	5.306	5.351	5.301	5.298	5.405	5.294	5.311	5.456	7.040
LassoLars	5.504	5.351	5.295	5.335	**5.271**	5.289	5.384	5.276	5.293	5.396	7.040
OLS	5.552	5.576	5.511	5.445	5.588	5.489	5.541	5.473	5.687	7.230	6.679
RF	6.177	6.091	6.428	6.177	6.131	6.486	6.486	6.603	6.095	6.241	7.519
SVR-linear-e	5.540	5.567	5.544	5.546	5.492	5.483	5.499	5.395	5.465	5.600	6.639
SVR-linear-e2	5.539	5.607	5.587	5.544	5.492	5.489	5.495	5.438	5.486	5.613	6.679
SVR-poly-2	5.947	6.079	6.267	5.967	6.143	6.255	6.577	6.826	6.321	8.308	7.030
SVR-poly-3	7.691	7.150	7.637	6.641	6.845	6.519	6.937	6.997	6.848	6.988	8.297
SVR-poly-4	8.650	8.930	8.972	7.820	8.449	7.683	8.386	8.379	8.707	9.135	8.997
SVR-rbf	5.494	5.594	5.621	5.454	5.605	5.576	5.536	5.583	5.577	5.725	7.048
SVR-sigmoid	7.764	6.632	7.978	7.764	6.777	7.480	7.480	7.630	6.689	6.617	7.003

Table 5.22: The best results (MAE) for individual estimators and variable sets. Each cell contains the best value over all possible dummies and the inclusion of weather data. The highest value for given estimator class is in green, for the particular selection of variables in blue, and the overall highest value is in **bold**.

5.2.1 Hypothesis: Neural network models are more accurate than classical regression models

The hypothesis of the ability of ANN to outperform classical models was not confirmed nor refuted — while both of the classical ANN performed worse than the OLS and the difference was significant by using DM test with 5% significance level for networks selected by MAE (viz fig. 5.1) and RMSE (viz fig. 5.2), one of the used ANNs performed better than the ridge regression (`KRR-linear`) when the main criterion was MAE. Furthermore, the SVR is sometimes considered to be a type of ANN [31, 188] in which case linear SVR with ϵ-insensitive loss outperformed the OLS and the KRR models for selection by MAE (significant difference by DM test at 5% significance level, viz fig. 5.1) and RMSE (the difference is not significant by DM test at 5% significance level fig. 5.2), it was the 5th best estimator, only the Bayesian Ridge regression (BR), the Elastic Net and its special case Lasso outperformed the SVR.

5.2.2 Hypothesis: Regression forests are able to perform similarly as other commonly used models

The RF builds a set of DTs and each of the DT outputs piecewise constant function, thus the approximation is quite crude, however, it is possible that a set of such estimators is able to predict the price relatively well despite the crude approximation of the individual DTs. While the `RF` estimator was not the best (in ranked 15th by both RMSE and MAE), it performed better than SVR with sigmoid and polynomial kernel of degree 3 and 4. However, since the `RF` was worse than the `OLS` estimator at 5% significance level for both RMSE and MAE, the hypothesis of similar performance has to be refuted.

5.3 Overall results

The overall results shows 20 best estimator by either MAE (table 5.23) or RMSE (table 5.24). The top 20 consists of only Lasso models (`Lasso` and `LassoLArs` or their generalization (`EN`). These estimators have several common characteristics — all of them are using dummy variables for different weekdays (`D`) and not the weekend dummies (`W`). Furthermore, 18 estimators from the top 20 by MAE and 17 estimators from the top 20 by RMSE use the dummies for individual months (`M`).

The yearly dummies (`Y`) are not as important as 15 estimators from the top 20 by MAE but only 9 estimators from the top 20 by RMSE use them. It seems that the `Y` dummies are usefull for miniziming the RMSE as they are more frequent in the top 20 and also when two estimators differs only by the inclusion of the `Y` dummies, the estimator with these dummies performs better (e.g. 2nd and 4th, 5th and 11th). Quite different trend in the `Y` dummies is when comparing the estimators by MAE — less than half of the top 20 estimators uses them but the estimators without them performs better (e.g. 1st and 3rd, 2nd and 4th, or 8th and 18th).

Also, all of the top 20 by MAE and 17 of the top 20 by RMSE use the weather data. The two cases where the weather data are not used are actually cases when the identical estimator but with the weather data is already in the rankings —

1st (with weather data) and 8th (without weather data), 4th (with weather data) and 15th (without weather data), and 6th (with weather data) and 17th (without weather data). This suggests that while the inclusion of weather data is beneficial (viz section 5.3.1), the gain is rather small.

rMAE	Estimator	Var.	W.	Test RMSE	Test MAE	rRMSE	rPareto
1	LassoLars	T2,7,14,28,M7,M30-DMY	Yes	7.3558	**5.2710**	5	1
2	LassoLars	T2,M7,30-DMY	Yes	7.3689	**5.2763**	13	2
3	LassoLars	T2,7,14,28,M7,M30-DM	Yes	7.3491	**5.2814**	3	1
4	LassoLars	T2,M7,30-DM	Yes	7.3465	**5.2829**	2	1
5	LassoLars	T2,7,M7-DMY	Yes	7.3533	**5.2887**	4	2
6	EN	T2,M7,30-DMY	Yes	7.3877	**5.2898**	39	3
7	LassoLars	T2-7-DMY	Yes	7.3798	**5.2931**	26	3
8	Lasso	T2,M7,30-DMY	Yes	7.3896	**5.2937**	42	4
9	LassoLars	T2,3,7-DMY	Yes	7.3687	**5.2946**	12	3
10	Lasso	T2,7,M7-DMY	Yes	7.3647	**5.2977**	10	3
11	LassoLars	T2,7,M7-DM	Yes	7.3363	**5.2980**	1	1
12	EN	T2,7,M7-DMY	Yes	7.3662	**5.2980**	11	4
13	LassoLars	T2,7,14,28,M7,M30-DY	Yes	7.3896	**5.3012**	43	5
14	Lasso	T2,7,14,28,M7,M30-DMY	Yes	7.3848	**5.3014**	35	5
15	EN	T2,7,14,28,M7,M30-DMY	Yes	7.3855	**5.3021**	38	6
16	LassoLars	T2,M7,30-DY	Yes	7.3923	**5.3023**	45	7
17	EN	T2,M7,30-DM	Yes	7.3839	**5.3051**	34	5
18	Lasso	T2,M7,30-DM	Yes	7.3838	**5.3052**	33	5
19	Lasso	T2,3,7-DMY	Yes	7.3785	**5.3057**	21	5
20	EN	T2,3,7-DMY	Yes	7.3794	**5.3071**	23	6

Table 5.23: The list of 20 best estimators by MAE. Column *rMAE* contain the overall rank by MAE (equivalety for *rRMSE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rRMSE	Estimator	Var.	W.	Test RMSE	Test MAE	rMAE	rPareto
1	LassoLars	T2,7,M7-DM	Yes	**7.3363**	5.2980	11	1
2	LassoLars	T2,M7,30-DM	Yes	**7.3465**	5.2829	4	1
3	LassoLars	T2,7,14,28,M7,M30-DM	Yes	**7.3491**	5.2814	3	1
4	LassoLars	T2,7,M7-DMY	Yes	**7.3533**	5.2887	5	2
5	LassoLars	T2,7,14,28,M7,M30-DMY	Yes	**7.3558**	5.2710	1	1
6	Lasso	T2,7,M7-DM	Yes	**7.3566**	5.3081	21	3
7	EN	T2,7,M7-DM	Yes	**7.3578**	5.3083	22	4
8	LassoLars	T2,7,M7-DM	No	**7.3596**	5.3405	49	5
9	LassoLars	T2,7-DM	Yes	**7.3613**	5.3367	43	5
10	Lasso	T2,7,M7-DMY	Yes	**7.3647**	5.2977	10	3
11	EN	T2,7,M7-DMY	Yes	**7.3662**	5.2980	12	4
12	LassoLars	T2,3,7-DMY	Yes	**7.3687**	5.2946	9	3
13	LassoLars	T2,M7,30-DMY	Yes	**7.3689**	5.2763	2	2
14	EN	T2,7-DMY	Yes	**7.3728**	5.3382	46	6
15	LassoLars	T2,7,M7-DMY	No	**7.3746**	5.3356	42	5
16	LassoLars	T2,7,M7-D	Yes	**7.3746**	5.3371	44	6
17	Lasso	T2,7,M7-DM	No	**7.3758**	5.3554	68	7
18	LassoLars	T2,3,7-DM	Yes	**7.3768**	5.3161	25	5
19	LassoLars	T2,7,M7-DY	Yes	**7.3780**	5.3203	30	6
20	LassoLars	T2,7,14,28,M7,M30-D	Yes	**7.3783**	5.3136	24	5

Table 5.24: The list of 20 best estimators by RMSE. Column *rRMSE* contain the overall rank by RMSE (equivalety for *rMAE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

5.3.1 Hypothesis: Models that use weather information are more accurate than models that do not

The inclusion of weather data to the model can be both beneficial and harmful. Beneficial because the electricity demand is influenced by weather (e.g. high temperature implies the need of cooling). However, the inclusion of such data might be also harmful from the practial point of view because it might lead to overfitting for certain estimators. To compare the benefits of the weather data, the estimators were paired. Each pair consists of two identical estimators — one included weather data in the model, the other did not. Then the Diebold–Mariano (DM) test is used for assessing whether the difference in performance between these two estimators is significant. The DM was used with one sided alternative hypothesis that the first estimator is better than the other. The results is deemed significant if the p–value is at least 5%. The results of the tests are shown in fig. 5.3 (MAE) and fig. 5.4 (RMSE) where each bar represents the number of pairs that dominated.

The inclusion of weather variables mostly led to deprecated performance for models `AB-DT`, `AB-LR`, `ANN48-0.5d-linear`, `OLS`, `RF` and kernel methods with sigmoid kernel and polynomial kernel of even degree. All of such models are particularly prone to overfitting as shown in figs. B.1 and B.2. Interestingly, the inclusion of historical weather data was beneficial for `SVR-poly-3` despite the overfitting to which this estimator is also prone and which was present (viz figs. B.1 and B.2). While the `ANN84-0.5d-1200e` had more neurons than the `ANN48-0.5d-linear` and thus should be more prone to overfitting as the capacity of such network is much larger, it was not the case because the `ANN84-0.5d-1200e` employed the regularization of the weights using L2 regularizer which, together with the quite high dropout, limited the overfitting.

The inclusion of weather variables made no significant differences or had mostly mixed effects for models `AB-DT` (for MAE only), `BR` (for RMSE only), `KRR-linear` (for RMSE only), `Lars`, `SVR-linear-e` (for RMSE only), `SVR-linear-e2` (for RMSE only), and `SVR-poly-4`. While the tests was mostly inconclusive for `Lars`, other estimators usually exhibited mixed effects when there were significant differences for between the estimators but not in consistent direction. For some estimators (e.g. `BR`) the inclusion of weather was beneficial when compared the model using DM test with power $p = 1$ (MAE), but the results were mixed when using DM test with power $p = 2$ (RMSE).

The inclusion of weather variables was beneficial for estimators `ANN84-0.5d-1200e`, `BR` (for MAE only), `EN`, `KRR-linear` (for MAE only), `Lasso`, `LassoLars`, `SVR-linear-e` (for MAE only), `SVR-linear-e2` (for MAE only), `SVR-poly-3`, and `SVR-sigmoid` . Most importantly, the inclusion of weather variables was beneficial for the top performing models such as `LassoLars` (1st by both MAE and RMSE) or `Lasso` or `EN`.

As discussed in section 5.3, most estimators from top 20 by both RMSE and MAE uses the weather data but the absolute gain is rather small. Thus the models have to be quite robust against overfitting for the gain not to be offset by the rise in error caused by overfitting.

Since the inclusion of historical weather data was beneficial for almost all of the top performing estimators and furthermore the gain was statistically significant at the 5% significance level for most estimators from the estimator classes with the highest performance, the hypothesis is considered confirmed even though the inclusion of weather data was not beneficial for all of the used estimators.

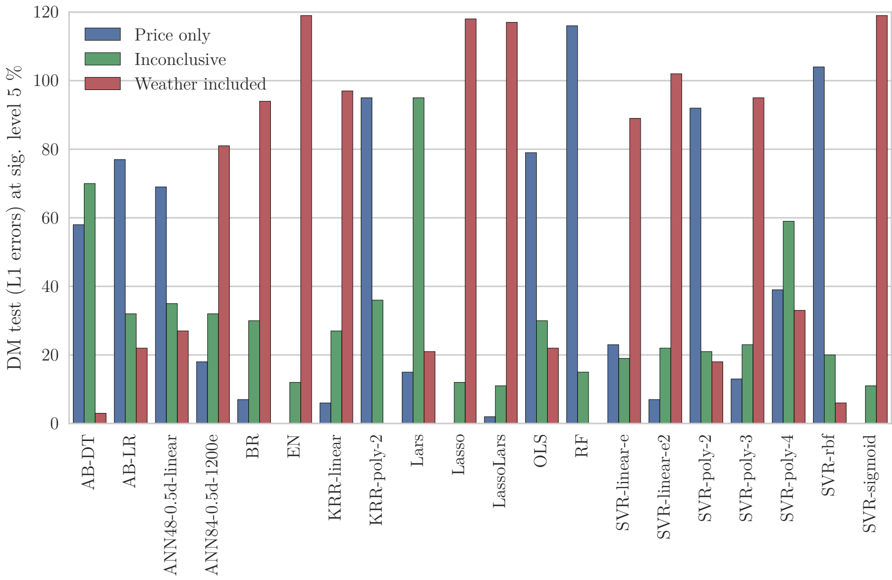


Figure 5.3: The plot shows the comparison of performance of identical estimator with and without the weather data. Each such pair of estimators was tested for significant dominance using the DM test with power $p = 1$ (MAE).

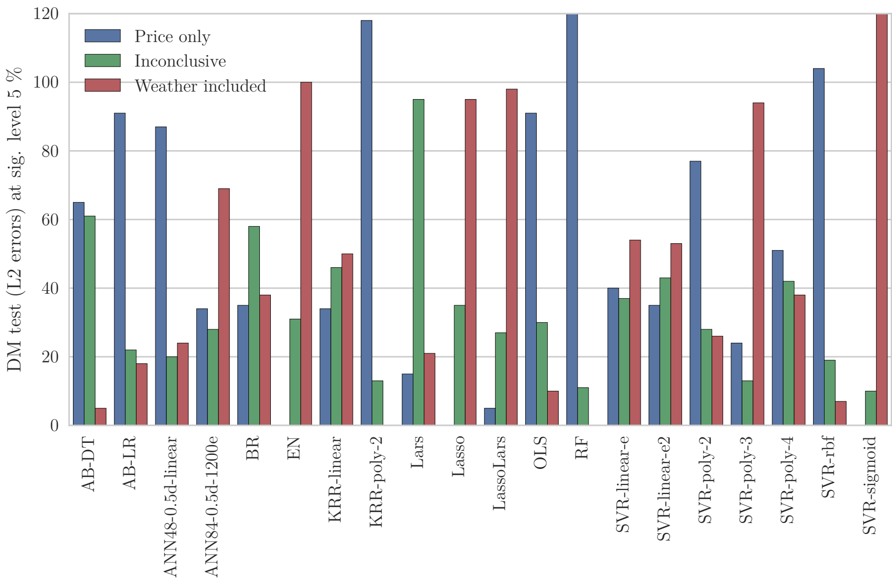


Figure 5.4: The plot shows the comparison of performance of identical estimator with and without the weather data. Each such pair of estimators was tested for significant dominance using the DM test with power $p = 2$ (RMSE).

Chapter 6

Conclusion

The electricity markets are very complex and will be even more complex in the future as the electricity generation mix is changing towards renewable resources, smardt grids are being introduced or are planned and as the end consumers of electriciy are starting to play more active roles. Therefore the access to good tools for analysis and prediction is necessary. There are many such tools for the stock market which, however, is significantly different from the electricity markets — e.g. the electricity cannot be stored economically, there are many seasonalities (daily, weekly, and yearly), and there fewer players on the electricity marekts. The goal of this thesis was to provide quite extensive comparison of possible model for electricity price prediction that could be used on the day ahead market (thus all the used models are forecasting day t using data only from days $t-2$ and before). Since it is impossible to cover all possible models, used estimators and their parametrizations, this work focuses on estimators using only the timeseries of historical electricity prices and historical weather data. Even with this narrow focus omitting some prominent classes of models (e.g. agent based models), the scope would be too large. Thus this work further narrows its focus only state-less regression models such as ridge regression, feed-forward neural networks or random forests. While some of the models maintaining internal states such as LSTM networks might be useful, the evaluation of such networks is available in my previous work [102] focusing on the prediction of electricity price volatility.

Despite the somewhat narrower focus, this work compares 20 different estimator classes with different variables which results in comparison of over 5000 different estimators. These include both estimators common in the literature focused on electricity price prediction such as artificial neural networks (ANNs), support vector regression (SVR), random forests (RFs) and OLS regression and estimators that are less frequent such as the AdaBoost, Lasso or Kernel ridge regression. Furthermore, since most of these estimators have parameters influencing their performance, the actual number of tested parametrizations of individual estimators is much higher as their parameters were optimized using the particle swarm optimization (PSO) and the Nelder–Mead (NM) method. The whole process took several weeks of processor time. To the extent of my knowledge, this is the most extensive comparison of models present in the literature — especially when compared to works focused on the Czech electricity market. While the compared models are not the only models that could be used, the presented comparison might provide good starting point for finding even better models.

A brief review of related work was provided in chapter 2. While thorough review would be out of the scope of this work, this chapter provided introduction into the used methods and good initial point for further research. The used methods were described in chapter 3. The description of methods PSO and NM methods used for optimization of estimators' parameters was provided in section 3.1, the individual estimators that were used in this thesis were described in section 3.2, and finally methods used for evaluation of the forecasts were introduced in section 3.3. Used data were described in chapter 4 — including the individual seasonalities present in the data. And finally, the estimators were compared in chapter 5. First, the estimators were evaluated within their estimator classes which assed the performance of estimators with different variables available (both price, weather and dummy variables) in section 5.1. This section also included the description of the used parameters for optimization of the individual estimators. Then the best estimators from each class were selected and used for comparison of different estimators classes in section 5.2. The best estimators were then tested using the DM tests for dominace — this test showed significant differences between most of the used estimator classes. Based on the results, the hypothesis on competitive performance of RFs was refuted. Furthermore, the tests have shown that the Lasso model optimized using modified Lars performs the best for all but one used lagged and mean variables. This made the `LassoLars` estimator the winner of this comparison and it is the recommended approach for electricity price prediction since the model itself is very simple and the estimator includes an implicit variable selection. Moreover this estimator is quite robust against overfitting and does not require lengthy training as ANNs estimators.

6.1 Future work

There is still much to be done in the comparison of possible approaches. The first possible extension is to broaden the scope of the used estimators to include, for example, the agent-based models. Another possible extension is to focus more on ANN and evaluate more than just two selected ANNs — the most viable possibility seems to use neuroevolution to search much greater space of possible network architectures and topologies including different regularization approaches that are available for ANNs. Since this search could be focused only on the best performing variable combinations from this thesis, the whole process would be less computationally costly because it wouldn't be necessary to evaluate the evolved network for more than 200 different variable combinations (in case of evolving separately topologies and weights). Yet another direction for possible extension of this work is to use historical weather forecasts instead of historical weather from day $t - 2$ and older. This could results in more accurate prediction as meteorological models are more accurate for short-term forecasts than just the regression of historical weather data. This work is limited also in the sense that it does not consider heterogenous ensemble estimators that consists of different estimators — the only ensemble estimators used in this work were random forests and AdaBoost. The future work should include also these estimators. Furthermore, predicting the prices is only part of the problem — another possible extension is to test the compared estimators in real market settings, e.g. use them in tools for trading the electricity. Also, the comparison should be repeated for different markets as it is possible that the differences between the markets will be quite large and different estimators might be more suitable for such markets than the `LassoLars` estimator that performed the best in the presented comparison.

Bibliography

[1] O. Abedinia, N. Amjady, M. Shafie-khah, and J. Catalão. "Electricity price forecast using Combinatorial Neural Network trained by a new stochastic search method". In: *Energy Conversion and Management* 105 (Nov. 2015), pp. 642–654. DOI: 10.1016/j.enconman.2015.08.025. URL: https://doi.org/10.1016/j.enconman.2015.08.025.

[2] S. K. Aggarwal, L. M. Saini, and A. Kumar. "Electricity price forecasting in deregulated markets: A review and evaluation". In: *International Journal of Electrical Power & Energy Systems* 31.1 (Jan. 2009), pp. 13–22. DOI: 10.1016/j.ijepes.2008.09.003. URL: https://doi.org/10.1016/j.ijepes.2008.09.003.

[3] M. A. H. Akhand, M. M. H. Rahman, and K. Murase. "Decision tree ensemble construction incorporating feature values modification and random subspace method". In: *2014 International Conference on Informatics, Electronics & Vision (ICIEV)*. Institute of Electrical & Electronics Engineers (IEEE), May 2014. DOI: 10.1109/iciev.2014.6850822. URL: http://dx.doi.org/10.1109/ICIEV.2014.6850822.

[4] M. Alamaniotis, A. Ikonomopoulos, D. Bargiotas, L. Tsoukalas, and A. Alamaniotis. "Day-ahead Electricity Price Forecasting using Optimized Multiple-Regression of Relevance Vector Machines". In: *8th Mediterranean Conference on Power Generation, Transmission, Distribution and Energy Conversion (MEDPOWER 2012)*. Institution of Engineering and Technology, 2012. DOI: 10.1049/cp.2012.2032. URL: https://doi.org/10.1049/cp.2012.2032.

[5] N. Amjady. "Day-Ahead Price Forecasting of Electricity Markets by a New Fuzzy Neural Network". In: *IEEE Transactions on Power Systems* 21.2 (May 2006), pp. 887–896. DOI: 10.1109/tpwrs.2006.873409. URL: https://doi.org/10.1109/tpwrs.2006.873409.

[6] N. Amjady, A. Daraeepour, and F. Keynia. "Day-ahead electricity price forecasting by modified relief algorithm and hybrid neural network". In: *IET Generation, Transmission & Distribution* 4.3 (2010), p. 432. DOI: 10.1049/iet-gtd.2009.0297. URL: https://doi.org/10.1049/iet-gtd.2009.0297.

[7] N. Amjady and M. Hemmati. "Energy price forecasting - problems and proposals for such predictions". In: *IEEE Power and Energy Magazine* 4.2 (Mar. 2006), pp. 20–29. DOI: 10.1109/mpae.2006.1597990. URL: http://dx.doi.org/10.1109/MPAE.2006.1597990.

[8] Anamika and N. Kumar. "Market Clearing Price prediction using ANN in Indian Electricity Markets". In: *2016 International Conference on Energy Efficient Technologies for Sustainability (ICEETS)*. IEEE, Apr. 2016. DOI: 10.1109/iceets.2016.7583797. URL: https://doi.org/10.1109/iceets.2016.7583797.

[9] P. Areekul, T. Senjyu, H. Toyama, and A. Yona. "Next day price forecasting for electricity market". In: *2011 International Conference on Advanced Power System Automation and Protection*. IEEE, Oct. 2011. DOI: 10.1109/apap.2011.6180594. URL: https://doi.org/10.1109/apap.2011.6180594.

[10] J. Armstrong and F. Collopy. "Error measures for generalizing about forecasting methods: Empirical comparisons". In: *International Journal of Forecasting* 8.1 (June 1992), pp. 69–80. DOI: 10.1016/0169-2070(92)90008-w. URL: https://doi.org/10.1016/0169-2070(92)90008-w.

[11] L. Auret and C. Aldrich. "Empirical comparison of tree ensemble variable importance measures". In: *Chemometrics and Intelligent Laboratory Systems* 105.2 (Feb. 2011), pp. 157–170. DOI: 10.1016/j.chemolab.2010.12.004. URL: http://dx.doi.org/10.1016/j.chemolab.2010.12.004.

[12] M. Bashari, A. Darudi, and N. Raeyatdoost. "Kalman Fusion algorithm in electricity price forecasting". In: *2014 14th International Conference on Environment and Electrical Engineering*. IEEE, May 2014. DOI: 10.1109/eeeic.2014.6835885. URL: https://doi.org/10.1109/eeeic.2014.6835885.

[13] J. Bastian, J. Zhu, V. Banunarayanan, and R. Mukerji. "Forecasting energy prices in a competitive market". In: *IEEE Computer Applications in Power* 12.3 (July 1999), pp. 40–45. DOI: 10.1109/67.773811. URL: https://doi.org/10.1109/67.773811.

[14] D. Benaouda and F. Murtagh. "Hybrid Wavelet Model for Electricity Pool-price Forecasting in a Deregulated Electricity Market". In: *2006 IEEE International Conference on Engineering of Intelligent Systems*. Institute of Electrical & Electronics Engineers (IEEE), 2006. DOI: 10.1109/iceis.2006.1703198. URL: http://dx.doi.org/10.1109/ICEIS.2006.1703198.

[15] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle. "Greedy Layer-wise Training of Deep Networks". In: *Proceedings of the 19th International Conference on Neural Information Processing Systems*. NIPS'06. Canada: MIT Press, 2006, pp. 153–160. URL: https://papers.nips.cc/paper/3048-greedy-layer-wise-training-of-deep-networks.pdf.

[16] J. Bergstra and Y. Bengio. "Random Search for Hyper-parameter Optimization". In: *J. Mach. Learn. Res.* 13.1 (Feb. 2012), pp. 281–305. ISSN: 1532-4435. URL: http://www.jmlr.org/papers/volume13/bergstra12a/bergstra12a.pdf.

[17] K. Bhaskar and S. Singh. "Wind power bidding strategy in a day-ahead electricity market". In: *2014 IEEE PES General Meeting | Conference & Exposition*. IEEE, July 2014. DOI: 10.1109/pesgm.2014.6939112. URL: https://doi.org/10.1109/pesgm.2014.6939112.

[18] R. Bisoi, O. Dharmapandit, and S. K. Dash. "A Functional link single layer feedforward neural network for electricity price forecasting using improved second order Levenberg-Marquardt algorithm". In: *2015 IEEE Power, Communication and Information Technology Conference (PCITC)*. IEEE, Oct. 2015. DOI: 10.1109/pcitc.2015.7438109. URL: https://doi.org/10.1109/pcitc.2015.7438109.

[19] M. R. Bonyadi and Z. Michalewicz. "Particle Swarm Optimization for Single Objective Continuous Space Problems: A Review". In: *Evolutionary Computation* 25.1 (Mar. 2017), pp. 1–54. DOI: 10.1162/evco_r_00180. URL: https://doi.org/10.1162/evco_r_00180.

[20] H. Bourlard and Y. Kamp. "Auto-association by multilayer perceptrons and singular value decomposition". In: *Biological Cybernetics* 59 (1988), pp. 291–294. URL: https://pdfs.semanticscholar.org/f582/1548720901c89b3b7481f7500d7cd64e99bd.pdf.

[21] L. Breiman. "Bagging predictors". In: *Machine Learning* 24.2 (1996), pp. 123–140. DOI: 10.1023/a:1018054314350. URL: http://dx.doi.org/10.1023/A:1018054314350.

[22] L. Breiman. "Random Forests". In: *Machine Learning* 45.1 (2001), pp. 5–32. DOI: 10.1023/a:1010933404324. URL: https://doi.org/10.1023/a:1010933404324.

[23] D. S. Broomhead and D. Lowe. "Radial basis functions, multi-variable functional interpolation and adaptive networks". In: *RSRE Memorandum No. 4148* (1988). URL: http://www.ling.upenn.edu/courses/cogs501/Rosenblatt1958.pdf.

[24] K. Černý. "Forecasting electricity prices in the Czech spot market". MA thesis. Charles University, Faculty of Social Science, Institute of Economic Studies, Feb. 2016. URL: https://is.cuni.cz/webapps/zzp/detail/147699.

[25] C.-C. Chang and C.-J. Lin. "LIBSVM: A library for support vector machines". In: *ACM Transactions on Intelligent Systems and Technology* 2 (3 2011). Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm, 27:1–27:27.

[26] F. Chollet et al. *Keras*. https://github.com/fchollet/keras. 2015.

[27] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio. *Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling*. 2014. eprint: arXiv:1412.3555. URL: https://arxiv.org/abs/1412.3555v1.

[28] M. Clerc and J. Kennedy. "The particle swarm - explosion, stability, and convergence in a multidimensional complex space". In: *IEEE Transactions on Evolutionary Computation* 6.1 (2002), pp. 58–73. DOI: 10.1109/4235.985692. URL: https://doi.org/10.1109/4235.985692.

[29] A. Conejo, M. Plazas, R. Espinola, and A. Molina. "Day-Ahead Electricity Price Forecasting Using the Wavelet Transform and ARIMA Models". In: *IEEE Transactions on Power Systems* 20.2 (May 2005), pp. 1035–1042. DOI: 10.1109/tpwrs.2005.846054. URL: https://doi.org/10.1109/tpwrs.2005.846054.

[30] J. Contreras, R. Espinola, F. Nogales, and A. Conejo. "ARIMA models to predict next-day electricity prices". In: *IEEE Transactions on Power Systems* 18.3 (Aug. 2003), pp. 1014–1020. DOI: 10.1109/tpwrs.2002.804943. URL: https://doi.org/10.1109/tpwrs.2002.804943.

[31] C. Cortes and V. Vapnik. "Support-vector networks". In: *Machine Learning* 20.3 (Sept. 1995), pp. 273–297. DOI: 10.1007/bf00994018. URL: https://doi.org/10.1007/bf00994018.

[32] J. C. Cuaresma, J. Hlouskova, S. Kossmeier, and M. Obersteiner. "Forecasting electricity spot-prices using linear univariate time-series models". In: *Applied Energy* 77.1 (Jan. 2004), pp. 87–106. DOI: 10.1016/s0306-2619(03)00096-5. URL: https://doi.org/10.1016/s0306-2619(03)00096-5.

[33] Y. N. Dauphin, H. de Vries, J. Chung, and Y. Bengio. "RMSProp and equilibrated adaptive learning rates for non-convex optimization". In: *CoRR* abs/1502.04390 (2015). URL: http://arxiv.org/abs/1502.04390.

[34] F. Davo, M. T. Vespucci, A. Gelmini, P. Grisi, and D. Ronzio. "Forecasting Italian electricity market prices using a Neural Network and a Support Vector Regression". In: *2016 AEIT International Annual Conference (AEIT)*. IEEE, 2016. DOI: 10.23919/aeit.2016.7892764. URL: https://doi.org/10.23919/aeit.2016.7892764.

[35] R. Deb, R. Albert, L.-L. Hsue, and N. Brown. "How to Incorporate Volatility and Risk in Electricity Price Forecasting". In: *The Electricity Journal* 13.4 (May 2000), pp. 65–75. DOI: 10.1016/s1040-6190(00)00102-0. URL: https://doi.org/10.1016/s1040-6190(00)00102-0.

[36] F. X. Diebold. "Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold–Mariano Tests". In: *Journal of Business & Economic Statistics* 33.1 (Jan. 2015), pp. 1–1. DOI: 10.1080/07350015.2014.983236. URL: https://doi.org/10.1080/07350015.2014.983236.

[37] F. X. Diebold and R. S. Mariano. "Comparing Predictive Accuracy". In: *Journal of Business & Economic Statistics* 13.1 (Jan. 1995), pp. 253–263. DOI: 10.1198/073500102753410444. URL: https://doi.org/10.1198/073500102753410444.

[38] T. G. Dietterich. "An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization". In: *Mach. Learn.* 40.2 (Aug. 2000), pp. 139–157. ISSN: 0885-6125. DOI: 10.1023/A:1007607513941. URL: http://dx.doi.org/10.1023/A:1007607513941.

[39] T. Dozat. *Incorporating Nesterov Momentum into Adam*. Tech. rep. 2015. URL: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.54.897.

[40] H. Drucker. "Improving Regressors Using Boosting Techniques". In: *Proceedings of the Fourteenth International Conference on Machine Learning*. ICML '97. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1997, pp. 107–115. ISBN: 1-55860-486-3. URL: http://dl.acm.org/citation.cfm?id=645526.657132.

[41] H. Drucker, C. J. C. Burges, L. Kaufman, A. J. Smola, and V. Vapnik. "Support Vector Regression Machines". In: *Advances in Neural Information Processing Systems 9*. Ed. by M. C. Mozer, M. I. Jordan, and T. Petsche. MIT Press, 1997, pp. 155–161. URL: http://papers.nips.cc/paper/1238-support-vector-regression-machines.pdf.

[42] J. Duchi, E. Hazan, and Y. Singer. "Adaptive Subgradient Methods for Online Learning and Stochastic Optimization". In: *J. Mach. Learn. Res.* 12 (July 2011), pp. 2121–2159. ISSN: 1532-4435. URL: http://www.jmlr.org/papers/volume12/duchi11a/duchi11a.pdf.

[43] E. E. Elattar. "Day-ahead price forecasting of electricity markets based on local informative vector machine". In: *IET Generation, Transmission & Distribution* 7.10 (Oct. 2013), pp. 1063–1071. DOI: 10.1049/iet-gtd.2012.0382. URL: https://doi.org/10.1049/iet-gtd.2012.0382.

[44] J. L. Elman. "Finding structure in time". In: *Cognitive science* 14.2 (1990), pp. 179–211. URL: https://crl.ucsd.edu/~elman/Papers/fsit.pdf.

[45] H. I. Elshazly, A. M. Elkorany, A. E. Hassanien, and A. T. Azar. "Ensemble classifiers for biomedical data: Performance evaluation". In: *2013 8th International Conference on Computer Engineering & Systems (ICCES)*. IEEE, Nov. 2013. DOI: 10.1109/icces.2013.6707198. URL: http://dx.doi.org/10.1109/ICCES.2013.6707198.

[46] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin. "LIBLINEAR: A Library for Large Linear Classification". In: *Journal of Machine Learning Research* 9 (2008), pp. 1871–1874.

[47] A. Fragkioudaki, A. Marinakis, and R. Cherkaoui. "Forecasting price spikes in European day-ahead electricity markets using decision trees". In: *2015 12th International Conference on the European Energy Market (EEM)*. IEEE, May 2015. DOI: 10.1109/eem.2015.7216672. URL: https://doi.org/10.1109/eem.2015.7216672.

[48] Y. Freund and R. E. Schapire. "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting". In: *Journal of Computer and System Sciences* 55.1 (Aug. 1997), pp. 119–139. DOI: 10.1006/jcss.1997.1504. URL: https://doi.org/10.1006/jcss.1997.1504.

[49] K. Fukushima. "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position". In: *Biological Cybernetics* 36.4 (Apr. 1980), pp. 193–202. DOI: 10.1007/bf00344251. URL: https://doi.org/10.1007%2Fbf00344251.

[50] F. Gao, X. Guan, X.-R. Cao, and A. Papalexopoulos. "Forecasting power market clearing price and quantity using a neural network method". In: *2000 Power Engineering Society Summer Meeting (Cat. No.00CH37134)*. IEEE, 2000. DOI: 10.1109/pess.2000.866984. URL: https://doi.org/10.1109/pess.2000.866984.

[51] R. Garcia, J. Contreras, M. vanAkkeren, and J. Garcia. "A GARCH Forecasting Model to Predict Day-Ahead Electricity Prices". In: *IEEE Transactions on Power Systems* 20.2 (May 2005), pp. 867–874. DOI: 10.1109/tpwrs.2005.846044. URL: https://doi.org/10.1109/tpwrs.2005.846044.

[52] C. García-Martos, J. Rodríguez, and M. J. Sánchez. "Forecasting electricity prices by extracting dynamic common factors: application to the Iberian Market". In: *IET Generation, Transmission & Distribution* 6.1 (2012), p. 11. DOI: 10.1049/iet-gtd.2011.0009. URL: https://doi.org/10.1049/iet-gtd.2011.0009.

[53] C. García-Martos, J. Rodríguez, and M. J. Sánchez. "Short-Term Forecasting of Electricity Prices Using Mixed Models". In: *Advances in Electric Power and Energy Systems*. John Wiley & Sons, Inc., June 2017, pp. 153–214. DOI: 10.1002/9781119260295.ch5. URL: https://doi.org/10.1002/9781119260295.ch5.

[54] R. Gareta, L. M. Romeo, and A. Gil. "Forecasting of electricity prices with neural networks". In: *Energy Conversion and Management* 47.13-14 (Aug. 2006), pp. 1770–1778. DOI: 10.1016/j.enconman.2005.10.010. URL: https://doi.org/10.1016/j.enconman.2005.10.010.

[55] R. Genuer, J.-M. Poggi, and C. Tuleau-Malot. "Variable selection using random forests". In: *Pattern Recognition Letters* 31.14 (Oct. 2010), pp. 2225–2236. DOI: 10.1016/j.patrec.2010.03.014. URL: http://dx.doi.org/10.1016/j.patrec.2010.03.014.

[56] P. S. Georgilakis. "Market Clearing Price Forecasting in Deregulated Electricity Markets Using Adaptively Trained Neural Networks". In: *Advances in Artificial Intelligence*. Springer Berlin Heidelberg, 2006, pp. 56–66. DOI: 10.1007/11752912_8. URL: https://doi.org/10.1007/11752912_8.

[57] C. González, I. Juárez, and J. Mira-McWilliams. "Important variable assessment and electricity price forecasting based on regression tree models: classification and regression trees, Bagging and Random Forests". In: *IET Generation, Transmission & Distribution* 9.11 (Aug. 2015), pp. 1120–1128. DOI: 10.1049/iet-gtd.2014.0655. URL: https://doi.org/10.1049/iet-gtd.2014.0655.

[58] V. Gonzalez, J. Contreras, and D. W. Bunn. "Forecasting prices and volatilities using a hybrid fundamental econometric model". In: *IEEE PES General Meeting*. IEEE, July 2010. DOI: 10.1109/pes.2010.5590100. URL: https://doi.org/10.1109/pes.2010.5590100.

[59] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. *Generative Adversarial Networks*. 2014. eprint: arXiv:1406.2661. URL: https://arxiv.org/pdf/1406.2661v1.pdf.

[60] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. http://www.deeplearningbook.org. MIT Press, 2016.

[61] A. Graves, G. Wayne, and I. Danihelka. *Neural Turing Machines*. 2014. eprint: arXiv:1410.5401. URL: https://arxiv.org/abs/1410.5401v2.

[62] L. Grossi and F. Nan. "Robust Self Exciting Threshold AutoRegressive models for electricity prices". In: *11th International Conference on the European Energy Market (EEM14)*. IEEE, May 2014. DOI: 10.1109/eem.2014.6861246. URL: https://doi.org/10.1109/eem.2014.6861246.

[63] L. Grossi and F. Nan. "Robust smooth transition threshold autoregressive models for electricity prices". In: *2015 12th International Conference on the European Energy Market (EEM)*. IEEE, May 2015. DOI: 10.1109/eem.2015.7216666. URL: https://doi.org/10.1109/eem.2015.7216666.

[64] J.-J. Guo and P. Luh. "Selecting input factors for clusters of gaussian radial basis function networks to improve market clearing price prediction". In: *IEEE Transactions on Power Systems* 18.2 (May 2003), pp. 665–672. DOI: 10.1109/tpwrs.2003.811012. URL: https://doi.org/10.1109/tpwrs.2003.811012.

[65] J.-J. Guo and P. Luh. "Improving Market Clearing Price Prediction by Using a Committee Machine of Neural Networks". In: *IEEE Transactions on Power Systems* 19.4 (Nov. 2004), pp. 1867–1876. DOI: 10.1109/tpwrs.2004.837759. URL: https://doi.org/10.1109/tpwrs.2004.837759.

[66] A. Hapfelmeier and K. Ulm. "A new variable selection approach using Random Forests". In: *Computational Statistics & Data Analysis* 60 (Apr. 2013), pp. 50–69. DOI: 10.1016/j.csda.2012.09.020. URL: http://dx.doi.org/10.1016/j.csda.2012.09.020.

[67] A. Hapfelmeier and K. Ulm. "Variable selection by Random Forests using data with missing values". In: *Computational Statistics & Data Analysis* 80 (Dec. 2014), pp. 129–139. DOI: 10.1016/j.csda.2014.06.017. URL: http://dx.doi.org/10.1016/j.csda.2014.06.017.

[68] K. He, X. Zhang, S. Ren, and J. Sun. "Deep Residual Learning for Image Recognition". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Institute of Electrical and Electronics Engineers (IEEE), June 2016. DOI: 10.1109/cvpr.2016.90. URL: https://doi.org/10.1109%2Fcvpr.2016.90.

[69] D. Hebb. *The Organization of Behavior*. New York: Wiley, 1949.

[70] G. E. Hinton and T. J. Sejnowski. "Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1". In: ed. by D. E. Rumelhart, J. L. McClelland, and C. PDP Research Group. Cambridge, MA, USA: MIT Press, 1986. Chap. Learning and Relearning in Boltzmann Machines, pp. 282–317. ISBN: 0-262-68053-X. URL: https://www.researchgate.net/publication/242509302_Learning_and_relearning_in_Boltzmann_machines.

[71] T. K. Ho. "The random subspace method for constructing decision forests". In: *IEEE Trans. Pattern Anal. Machine Intell.* 20.8 (1998), pp. 832–844. DOI: 10.1109/34.709601. URL: http://dx.doi.org/10.1109/34.709601.

[72] S. Hochreiter and J. Schmidhuber. "Long Short-Term Memory". In: *Neural Computation* 9.8 (Nov. 1997), pp. 1735–1780. DOI: 10.1162/neco.1997.9.8.1735. URL: https://doi.org/10.1162%2Fneco.1997.9.8.1735.

[73] A. E. Hoerl and R. W. Kennard. "Ridge Regression: Biased Estimation for Nonorthogonal Problems". In: *Technometrics* 12.1 (Feb. 1970), pp. 55–67. DOI: 10.1080/00401706.1970.10488634. URL: https://doi.org/10.1080/00401706.1970.10488634.

[74] S. Honari, J. Yosinski, P. Vincent, and C. Pal. "Recombinator Networks: Learning Coarse-to-Fine Feature Aggregation". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Institute of Electrical and Electronics Engineers (IEEE), June 2016. DOI: 10.1109/cvpr.2016.619. URL: https://doi.org/10.1109%2Fcvpr.2016.619.

[75] Y.-Y. Hong and C.-F. Lee. "A neuro-fuzzy price forecasting approach in deregulated electricity markets". In: *Electric Power Systems Research* 73.2 (Feb. 2005), pp. 151–157. DOI: 10.1016/j.epsr.2004.07.002. URL: https://doi.org/10.1016/j.epsr.2004.07.002.

[76] Y.-Y. Hong and C.-Y. Hsiao. "Locational marginal price forecasting in deregulated electricity markets using artificial intelligence". In: *IEE Proceedings - Generation, Transmission and Distribution* 149.5 (2002), p. 621. DOI: 10.1049/ip-gtd:20020371. URL: https://doi.org/10.1049/ip-gtd:20020371.

[77] J. J. Hopfield. "Neural networks and physical systems with emergent collective computational abilities". In: *Proceedings of the National Academy of Sciences of the United States of America*. Vol. 79. 8. 1982, pp. 2554–2558. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC346238/.

[78] Z. Hu, L. Yang, Z. Wang, D. Gan, W. Sun, and K. Wang. "A game-theoretic model for electricity markets with tight capacity constraints". In: *International Journal of Electrical Power & Energy Systems* 30.3 (Mar. 2008), pp. 207–215. DOI: 10.1016/j.ijepes.2007.07.008. URL: https://doi.org/10.1016/j.ijepes.2007.07.008.

[79] Z. Hu, Y. Yu, Z. Wang, W. Sun, D. Gan, and Z. Han. "Price forecasting using an integrated approach". In: *2004 IEEE International Conference on Electric Utility Deregulation, Restructuring and Power Technologies. Proceedings.* IEEE, 2004. DOI: 10.1109/drpt.2004.1338463. URL: https://doi.org/10.1109/drpt.2004.1338463.

[80] G. Huang, S. Song, C. Wu, and K. You. "Robust Support Vector Regression for Uncertain Input and Output Data". In: *IEEE Transactions on Neural Networks and Learning Systems* 23.11 (Nov. 2012), pp. 1690–1700. DOI: 10.1109/tnnls.2012.2212456. URL: https://doi.org/10.1109/tnnls.2012.2212456.

[81] R. J. Hyndman and A. B. Koehler. "Another look at measures of forecast accuracy". In: *International Journal of Forecasting* 22.4 (Oct. 2006), pp. 679–688. DOI: 10.1016/j.ijforecast.2006.03.001. URL: https://doi.org/10.1016/j.ijforecast.2006.03.001.

[82] H. Jaeger. "Harnessing Nonlinearity: Predicting Chaotic Systems and Saving Energy in Wireless Communication". In: *Science* 304.5667 (Apr. 2004), pp. 78–80. DOI: 10.1126/science.1091277. URL: https://doi.org/10.1126%2Fscience.1091277.

[83] T. Jonsson, P. Pinson, H. A. Nielsen, H. Madsen, and T. S. Nielsen. "Forecasting Electricity Spot Prices Accounting for Wind Power Predictions". In: *IEEE Transactions on Sustainable Energy* 4.1 (Jan. 2013), pp. 210–218. DOI: 10.1109/tste.2012.2212731. URL: https://doi.org/10.1109/tste.2012.2212731.

[84] T. Kanamori, S. Fujiwara, and A. Takeda. "Breakdown Point of Robust Support Vector Machines". In: *Entropy* 19.2 (Feb. 2017), p. 83. DOI: 10.3390/e19020083. URL: https://doi.org/10.3390/e19020083.

[85] A. Karpathy. *CS231n: Convolutional Neural Networks for Visual Recognition — Module 2: Convolutional Neural Networks*. 2017. URL: http://cs231n.github.io/neural-networks-2/ (visited on 05/09/2017).

[86] V. Kekatos, Y. Zhang, and G. B. Giannakis. "Electricity Market Forecasting via Low-Rank Multi-Kernel Learning". In: *IEEE Journal of Selected Topics in Signal Processing* 8.6 (Dec. 2014), pp. 1182–1193. DOI: 10.1109/jstsp.2014.2336611. URL: https://doi.org/10.1109/jstsp.2014.2336611.

[87] V. Kekatos, Y. Zhang, and G. B. Giannakis. "Low-rank kernel learning for electricity market inference". In: *2013 Asilomar Conference on Signals, Systems and Computers*. IEEE, Nov. 2013. DOI: 10.1109/acssc.2013.6810605. URL: https://doi.org/10.1109/acssc.2013.6810605.

[88] A. V. Kelarev, A. Stranieri, J. L. Yearwood, and H. F. Jelinek. "Empirical Study of Decision Trees and Ensemble Classifiers for Monitoring of Diabetes Patients in Pervasive Healthcare". In: *2012 15th International Conference on Network-Based Information Systems*. IEEE, Sept. 2012. DOI: 10.1109/nbis.2012.20. URL: http://dx.doi.org/10.1109/NBiS.2012.20.

[89] J. Kennedy and R. Eberhart. "Particle swarm optimization". In: *Proceedings of ICNN'95 - International Conference on Neural Networks*. IEEE, 1995. DOI: 10.1109/icnn.1995.488968. URL: https://doi.org/10.1109/icnn.1995.488968.

[90] A. Kian and A. Keyhani. "Stochastic price modeling of electricity in deregulated energy markets". In: *Proceedings of the 34th Annual Hawaii International Conference on System Sciences*. IEEE Comput. Soc, 2001. DOI: 10.1109/hicss.2001.926292. URL: https://doi.org/10.1109/hicss.2001.926292.

[91] C.-i. Kim, I.-K. Yu, and Y. Song. "Prediction of system marginal price of electricity using wavelet transform analysis". In: *Energy Conversion and Management* 43.14 (Sept. 2002), pp. 1839–1851. DOI: 10.1016/s0196-8904(01)00127-3. URL: https://doi.org/10.1016/s0196-8904(01)00127-3.

[92] S. Kim and H. Kim. "A new metric of absolute percentage error for intermittent demand forecasts". In: *International Journal of Forecasting* 32.3 (July 2016), pp. 669–679. DOI: 10.1016/j.ijforecast.2015.12.003. URL: https://doi.org/10.1016/j.ijforecast.2015.12.003.

[93] D. P. Kingma and J. Ba. "Adam: A Method for Stochastic Optimization". In: *CoRR* abs/1412.6980 (2014). URL: http://arxiv.org/abs/1412.6980.

[94] D. P. Kingma and M. Welling. *Auto-association by multilayer perceptrons and singular value decomposition*. arXiv preprint. 2013. URL: https://arxiv.org/abs/1312.6114.

[95] A. M. Kintsakis, A. Chrysopoulos, and P. A. Mitkas. "Agent-based short-term load and price forecasting using a parallel implementation of an adaptive PSO-trained local linear wavelet neural network". In: *2015 12th International Conference on the European Energy Market (EEM)*. IEEE, May 2015. DOI: 10.1109/eem.2015.7216611. URL: https://doi.org/10.1109/eem.2015.7216611.

[96] G. Klaboe. "Forecasting hourly electricity prices for bidding optimization". In: *2015 12th International Conference on the European Energy Market (EEM)*. IEEE, May 2015. DOI: 10.1109/eem.2015.7216636. URL: https://doi.org/10.1109/eem.2015.7216636.

[97] T. Kohonen. "Self-organized formation of topologically correct feature maps". In: *Biological Cybernetics* 43.1 (1982), pp. 59–69. DOI: 10.1007/bf00337288. URL: https://doi.org/10.1007%2Fbf00337288.

[98] S. B. Kotsiantis. "Decision trees: a recent overview". In: *Artificial Intelligence Review* 39.4 (June 2011), pp. 261–283. DOI: 10.1007/s10462-011-9272-4. URL: http://dx.doi.org/10.1007/s10462-011-9272-4.

[99] D. Kotur and M. Zarkovic. "Neural network models for electricity prices and loads short and long-term prediction". In: *2016 4th International Symposium on Environmental Friendly Energies and Applications (EFEA)*. IEEE, Sept. 2016. DOI: 10.1109/efea.2016.7748787. URL: https://doi.org/10.1109/efea.2016.7748787.

[100] Š. Krška. "Current Situation on the Czech Electricity Market: with an Emphasis on the Fourth Regulatory Period of the Czech Energy Regulatory Office". MA thesis. Charles University, Faculty of Social Science, Institute of Economic Studies, June 2014. URL: https://is.cuni.cz/webapps/zzp/detail/133785.

[101] T. D. Kulkarni, W. F. Whitney, P. Kohli, and J. Tenenbaum. "Deep Convolutional Inverse Graphics Network". In: *Advances in Neural Information Processing Systems 28*. Ed. by C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett. Curran Associates, Inc., 2015, pp. 2539–2547. URL: http://papers.nips.cc/paper/5851-deep-convolutional-inverse-graphics-network.pdf.

[102] Kunc. "Electricity market: Analysis and prediction of volatility". Charles University, Faculty of Social Science, Institute of Economic Studies, July 2015. URL: https://is.cuni.cz/webapps/zzp/detail/151474/.

[103] L. I. Kuncheva. *Combining Pattern Classifiers*. John Wiley & Sons, Inc., July 2004. DOI: 10.1002/0471660264. URL: http://dx.doi.org/10.1002/0471660264.

[104] L. I. Kuncheva and J. J. Rodríguez. "An Experimental Study on Rotation Forest Ensembles". In: *Multiple Classifier Systems*. Springer Science + Business Media, 2007, pp. 459–468. DOI: 10.1007/978-3-540-72523-7_46. URL: http://dx.doi.org/10.1007/978-3-540-72523-7_46.

[105] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright. "Convergence Properties of the Nelder–Mead Simplex Method in Low Dimensions". In: *SIAM Journal on Optimization* 9.1 (Jan. 1998), pp. 112–147. DOI: 10.1137/s1052623496303470. URL: https://doi.org/10.1137/s1052623496303470.

[106] A. Lahouar and J. B. H. Slama. "Comparative study of learning machine predictors for half-hour and day-ahead electricity price forecast in deregulated markets". In: *2016 7th International Renewable Energy Congress (IREC)*. IEEE, Mar. 2016. DOI: 10.1109/irec.2016.7478870. URL: https://doi.org/10.1109/irec.2016.7478870.

[107] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE 86.11*. 1998, pp. 2278–2324. URL: http://yann.lecun.com/exdb/publis/pdf/lecun-98.pdf.

[108] M. Lei and Z. Feng. "Modeling Electricity Price Forecast with Grey and Correlation Method in Competitive Markets". In: *2010 Asia-Pacific Power and Energy Engineering Conference*. IEEE, 2010. DOI: 10.1109/appeec.2010.5448701. URL: https://doi.org/10.1109/appeec.2010.5448701.

[109] C. Li and S. Wang. "Next-Day Power Market Clearing Price Forecasting Using Artificial Fish-Swarm Based Neural Network". In: *Advances in Neural Networks - ISNN 2006*. Springer Berlin Heidelberg, 2006, pp. 1290–1295. DOI: 10.1007/11760023_187. URL: https://doi.org/10.1007/11760023_187.

[110] K. Li and Y. Han. "Study of selective ensemble learning method and its diversity based on decision tree and neural network". In: *2010 Chinese Control and Decision Conference*. Institute of Electrical & Electronics Engineers (IEEE), May 2010. DOI: 10.1109/ccdc.2010.5498179. URL: http://dx.doi.org/10.1109/CCDC.2010.5498179.

[111] Z. Lin, L. Gao, and D. Zhang. "Predictions of System Marginal Price of Electricity Using Recurrent Neural Network". In: *2006 6th World Congress on Intelligent Control and Automation*. IEEE, 2006. DOI: 10.1109/wcica.2006.1713442. URL: https://doi.org/10.1109/wcica.2006.1713442.

[112] S. Linnainmaa. "Taylor expansion of the accumulated rounding error". In: *BIT* 16.2 (June 1976), pp. 146–160. DOI: 10.1007/bf01931367. URL: https://doi.org/10.1007%2Fbf01931367.

[113] W.-Y. Loh. "Classification and regression trees". In: *WIREs Data Mining Knowl Discov* 1.1 (Jan. 2011), pp. 14–23. DOI: 10.1002/widm.8. URL: http://dx.doi.org/10.1002/widm.8.

[114] W.-Y. Loh. “Fifty Years of Classification and Regression Trees”. In: *International Statistical Review* 82.3 (June 2014), pp. 329–348. DOI: 10.1111/insr.12016. URL: https://doi.org/10.1111/insr.12016.

[115] D. Lopez, J. Juan, and J. Carpio. “EM estimation of multivariate dynamic models for predicting electricity prices”. In: *2010 7th International Conference on the European Energy Market*. IEEE, June 2010. DOI: 10.1109/eem.2010.5558693. URL: https://doi.org/10.1109/eem.2010.5558693.

[116] A. T. Lora, J. M. R. Santos, A. G. Exposito, J. L. M. Ramos, and J. C. R. Santos. “Electricity Market Price Forecasting Based on Weighted Nearest Neighbors Techniques”. In: *IEEE Transactions on Power Systems* 22.3 (Aug. 2007), pp. 1294–1301. DOI: 10.1109/tpwrs.2007.901670. URL: https://doi.org/10.1109/tpwrs.2007.901670.

[117] A. T. Lora, J. R. Santos, J. R. Santos, A. G. Expósito, and J. L. M. Ramos. “A Comparison of Two Techniques for Next-Day Electricity Price Forecasting”. In: *Intelligent Data Engineering and Automated Learning — IDEAL 2002*. Springer Berlin Heidelberg, 2002, pp. 384–390. DOI: 10.1007/3-540-45675-9_57. URL: https://doi.org/10.1007/3-540-45675-9_57.

[118] X. Lu, Z. Dong, and X. Li. “Electricity market price spike forecast with data mining techniques”. In: *Electric Power Systems Research* 73.1 (Jan. 2005), pp. 19–29. DOI: 10.1016/s0378-7796(04)00125-7. URL: https://doi.org/10.1016/s0378-7796(04)00125-7.

[119] Y. Ma, P. Luh, K. Kasiviswanathan, and E. Ni. “A neural network-based method for forecasting zonal locational marginal prices”. In: *IEEE Power Engineering Society General Meeting*. IEEE, 2004. DOI: 10.1109/pes.2004.1372802. URL: https://doi.org/10.1109/pes.2004.1372802.

[120] W. Maass, T. Natschläger, and H. Markram. “Real-Time Computing Without Stable States: A New Framework for Neural Computation Based on Perturbations”. In: *Neural Computation* 14.11 (Nov. 2002), pp. 2531–2560. DOI: 10.1162/089976602760407955. URL: https://doi.org/10.1162%2F089976602760407955.

[121] M. Maca and L. Lhotska. “Wrapper Feature Selection Significantly Improves Nonlinear Prediction of Electricity Spot Prices”. In: *2013 IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, Oct. 2013. DOI: 10.1109/smc.2013.203. URL: https://doi.org/10.1109/smc.2013.203.

[122] D. Masri, H. Zeineldin, and W. L. Woon. “Electricity price and demand forecasting under smart grid environment”. In: *2015 IEEE 15th International Conference on Environment and Electrical Engineering (EEEIC)*. IEEE, June 2015. DOI: 10.1109/eeeic.2015.7165472. URL: https://doi.org/10.1109/eeeic.2015.7165472.

[123] A. MateoGonzalez, A. MunozSanRoque, and J. Garcia-Gonzalez. “Modeling and Forecasting Electricity Prices with Input/Output Hidden Markov Models”. In: *IEEE Transactions on Power Systems* 20.1 (Feb. 2005), pp. 13–24. DOI: 10.1109/tpwrs.2004.840412. URL: https://doi.org/10.1109/tpwrs.2004.840412.

[124] D. H. Mazengia and L. A. Tuan. “Forecasting spot electricity market prices using time series models”. In: *2008 IEEE International Conference on Sustainable Energy Technologies*. IEEE, Nov. 2008. DOI: 10.1109/icset.2008.4747199. URL: https://doi.org/10.1109/icset.2008.4747199.

[125] W. S. McCulloch and W. Pitts. “A logical calculus of the ideas immanent in nervous activity”. In: *The Bulletin of Mathematical Biophysics* 5.4 (Dec. 1943), pp. 115–133. DOI: 10.1007/bf02478259. URL: https://doi.org/10.1007%2Fbf02478259.

[126] K. I. M. McKinnon. “Convergence of the Nelder–Mead Simplex Method to a Nonstationary Point”. In: *SIAM Journal on Optimization* 9.1 (Jan. 1998), pp. 148–158. DOI: 10.1137/s1052623496303482. URL: https://doi.org/10.1137/s1052623496303482.

[127] M. R. V. Moghadam, K. Afshar, and N. Bigdeli. “A New Hybrid Method for Short-Term Price Forecasting in Iran Electricity Market”. In: *2011 International Conference on Management and Service Science*. IEEE, Aug. 2011. DOI: 10.1109/icmss.2011.5998135. URL: https://doi.org/10.1109/icmss.2011.5998135.

[128] A. Mohamed and M. E. El-Hawary. “Effective input features selection for electricity price forecasting”. In: *2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*. IEEE, May 2016. DOI: 10.1109/ccece.2016.7726626. URL: https://doi.org/10.1109/ccece.2016.7726626.

[129] A. Mohamed and M. E. El-Hawary. “Mid-term electricity price forecasting using SVM”. In: *2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*. IEEE, May 2016. DOI: 10.1109/ccece.2016.7726765. URL: https://doi.org/10.1109/ccece.2016.7726765.

[130] A. Mohamed and M. E. El-Hawary. “On optimization of SVMs kernels and parameters for electricity price forecasting”. In: *2016 IEEE Electrical Power and Energy Conference (EPEC)*. IEEE, Oct. 2016. DOI: 10.1109/epec.2016.7771773. URL: https://doi.org/10.1109/epec.2016.7771773.

[131] F. Moscetti, S. Paoletti, and A. Vicino. “Analysis and models of electricity prices in the Italian ancillary services market”. In: *IEEE PES Innovative Smart Grid Technologies, Europe*. IEEE, Oct. 2014. DOI: 10.1109/isgteurope.2014.7028755. URL: https://doi.org/10.1109/isgteurope.2014.7028755.

[132] A. Motamedi, H. Zareipour, and W. D. Rosehart. “Electricity Price and Demand Forecasting in Smart Grids”. In: *IEEE Transactions on Smart Grid* 3.2 (June 2012), pp. 664–674. DOI: 10.1109/tsg.2011.2171046. URL: https://doi.org/10.1109/tsg.2011.2171046.

[133] A. Navada, A. N. Ansari, S. Patil, and B. A. Sonkamble. “Overview of use of decision tree algorithms in machine learning”. In: *2011 IEEE Control and System Graduate Research Colloquium*. IEEE, June 2011. DOI: 10.1109/icsgrc.2011.5991826. URL: http://dx.doi.org/10.1109/ICSGRC.2011.5991826.

[134] J. A. Nelder and R. Mead. “A Simplex Method for Function Minimization”. In: *The Computer Journal* 7.4 (Jan. 1965), pp. 308–313. DOI: 10.1093/comjnl/7.4.308. URL: https://doi.org/10.1093/comjnl/7.4.308.

[135] E. Ni and P. Luh. “Forecasting power market clearing price and its discrete PDF using a Bayesian-based classification method”. In: *2001 IEEE Power Engineering Society Winter Meeting. Conference Proceedings (Cat. No.01CH37194)*. IEEE, 2001. DOI: 10.1109/pesw.2001.917338. URL: https://doi.org/10.1109/pesw.2001.917338.

[136] J. Nicolaisen, C. Richter, and G. Sheble. “Price signal analysis for competitive electric generation companies”. In: *DRPT2000. International Conference on Electric Utility Deregulation and Restructuring and Power Technologies. Proceedings (Cat. No.00EX382)*. IEEE, 2000. DOI: 10.1109/drpt.2000.855640. URL: https://doi.org/10.1109/drpt.2000.855640.

[137] T. Niimura, H.-S. Ko, and K. Ozawa. “A day-ahead electricity price prediction based on a fuzzy-neuro autoregressive model in a deregulated electricity market”. In: *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No.02CH37290)*. IEEE, 2002. DOI: 10.1109/ijcnn.2002.1007714. URL: https://doi.org/10.1109/ijcnn.2002.1007714.

[138] F. Nogales, J. Contreras, A. Conejo, and R. Espinola. “Forecasting next-day electricity prices by time series models”. In: *IEEE Transactions on Power Systems* 17.2 (May 2002), pp. 342–348. DOI: 10.1109/tpwrs.2002.1007902. URL: http://dx.doi.org/10.1109/TPWRS.2002.1007902.

[139] C. Nunes, A. Pacheco, and T. Silva. “Statistical models to predict electricity prices”. In: *2008 5th International Conference on the European Electricity Market*. IEEE, May 2008. DOI: 10.1109/eem.2008.4579004. URL: https://doi.org/10.1109/eem.2008.4579004.

[140] G. J. O. Osorio, J. C. O. Matias, and J. P. S. Catalao. “Hybrid evolutionary-adaptive approach to predict electricity prices and wind power in the short-term”. In: *2014 Power Systems Computation Conference*. IEEE, Aug. 2014. DOI: 10.1109/pscc.2014.7038453. URL: https://doi.org/10.1109/pscc.2014.7038453.

[141] I. P. Panapakidis and A. S. Dagoumas. “Day-ahead electricity price forecasting via the application of artificial neural network based models”. In: *Applied Energy* 172 (June 2016), pp. 132–151. DOI: 10.1016/j.apenergy.2016.03.089. URL: https://doi.org/10.1016/j.apenergy.2016.03.089.

[142] H.-T. Pao. “A Neural Network Approach to m-Daily-Ahead Electricity Price Prediction”. In: *Advances in Neural Networks - ISNN 2006*. Springer Berlin Heidelberg, 2006, pp. 1284–1289. DOI: 10.1007/11760023_186. URL: https://doi.org/10.1007/11760023_186.

[143] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830. URL: http://arxiv.org/pdf/1201.0490v2.pdf.

[144] R. Pereira, T. M. Sousa, T. Pinto, I. Praça, Z. Vale, and H. Morais. "Strategic Bidding for Electricity Markets Negotiation Using Support Vector Machines". In: *Advances in Intelligent Systems and Computing*. Springer Nature, 2014, pp. 9–17. DOI: 10.1007/978-3-319-07476-4_2. URL: http://dx.doi.org/10.1007/978-3-319-07476-4_2.

[145] T. Pinto, T. M. Sousa, H. Morais, I. Praça, and Z. Vale. "Metalearning to support competitive electricity market players' strategic bidding". In: *Electric Power Systems Research* 135 (June 2016), pp. 27–34. DOI: 10.1016/j.epsr.2016.03.012. URL: http://dx.doi.org/10.1016/j.epsr.2016.03.012.

[146] R. Poli. "Analysis of the Publications on the Applications of Particle Swarm Optimisation". In: *Journal of Artificial Evolution and Applications* 2008 (2008), pp. 1–10. DOI: 10.1155/2008/685175. URL: https://doi.org/10.1155/2008/685175.

[147] H. Pousinho, V. Mendes, and J. Catalao. "Wavelet-neuro-fuzzy approach for predicting short-term electricity prices in a competitive market". In: *7th Mediterranean Conference and Exhibition on Power Generation, Transmission, Distribution and Energy Conversion (MedPower 2010)*. IET, 2010. DOI: 10.1049/cp.2010.0849. URL: https://doi.org/10.1049/cp.2010.0849.

[148] H. Qian and Z. Qiu. "Feature selection using C4.5 algorithm for electricity price prediction". In: *2014 International Conference on Machine Learning and Cybernetics*. IEEE, July 2014. DOI: 10.1109/icmlc.2014.7009113. URL: https://doi.org/10.1109/icmlc.2014.7009113.

[149] Z.-W. Qiu. "Mutivariable mutual information based feature selection for electricity price forecasting". In: *2012 International Conference on Machine Learning and Cybernetics*. IEEE, July 2012. DOI: 10.1109/icmlc.2012.6358906. URL: https://doi.org/10.1109/icmlc.2012.6358906.

[150] J. R. Quinlan. "Induction of decision trees". In: *Machine Learning* 1.1 (Mar. 1986), pp. 81–106. DOI: 10.1007/bf00116251. URL: http://dx.doi.org/10.1007/BF00116251.

[151] M. Rafiei, T. Niknam, and M.-H. Khooban. "Probabilistic Forecasting of Hourly Electricity Price by Generalization of ELM for Usage in Improved Wavelet Neural Network". In: *IEEE Transactions on Industrial Informatics* 13.1 (Feb. 2017), pp. 71–79. DOI: 10.1109/tii.2016.2585378. URL: https://doi.org/10.1109/tii.2016.2585378.

[152] M. Ranzato, C. Poultney, S. Chopra, and Y. LeCun. "Efficient learning of sparse representations with an energy-based model". In: *Proceedings of 20th International Conference on Neural Information Processing Systems*. 2007. URL: https://papers.nips.cc/paper/3112-efficient-learning-of-sparse-representations-with-an-energy-based-model.pdf.

[153] I. A. W. A. Razak, I. Z. Abidin, K. S. Yap, A. A. Z. Abidin, T. K. A. Rahman, and M. N. M. Nasir. "A novel hybrid method of LSSVM-GA with multiple stage optimization for electricity price forecasting". In: *2016 IEEE International Conference on Power and Energy (PECon)*. IEEE, Nov. 2016. DOI: 10.1109/pecon.2016.7951593. URL: https://doi.org/10.1109/pecon.2016.7951593.

[154] G. Ritschard. "CHAID and Earlier Supervised Tree Methods". In: *Cahiers du Département d'économétrie*. 2010, pp. 3–6. URL: http://www.unige.ch/ses/metri/cahiers/2010_02.pdf.

[155] C. Rodriguez and G. Anders. "Energy Price Forecasting in the Ontario Competitive Power System Market". In: *IEEE Transactions on Power Systems* 19.1 (Feb. 2004), pp. 366–374. DOI: 10.1109/tpwrs.2003.821470. URL: https://doi.org/10.1109/tpwrs.2003.821470.

[156] J. J. Rodriguez, L. I. Kuncheva, and C. J. Alonso. "Rotation Forest: A New Classifier Ensemble Method". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28.10 (Oct. 2006), pp. 1619–1630. DOI: 10.1109/tpami.2006.211. URL: http://dx.doi.org/10.1109/TPAMI.2006.211.

[157] L. Rokach. "Ensemble-based Classifiers". In: *Artif. Intell. Rev.* 33.1-2 (Feb. 2010), pp. 1–39. ISSN: 0269-2821. DOI: 10.1007/s10462-009-9124-7. URL: http://dx.doi.org/10.1007/s10462-009-9124-7.

[158] L. Rokach. "Taxonomy for characterizing ensemble methods in classification tasks: A review and annotated bibliography". In: *Computational Statistics & Data Analysis* 53.12 (Oct. 2009), pp. 4046–4072. DOI: 10.1016/j.csda.2009.07.017. URL: http://dx.doi.org/10.1016/j.csda.2009.07.017.

[159] F. Rosenblatt. "The Perceptron: A Probabilistic Model for Information Storage and Organization in The Brain". In: *Psychological Review* 65.6 (1958), pp. 386–408. URL: http://www.ling.upenn.edu/courses/cogs501/Rosenblatt1958.pdf.

[160] L. Rutkowski, M. Jaworski, L. Pietruczuk, and P. Duda. "The CART decision tree for mining data streams". In: *Information Sciences* 266 (May 2014), pp. 1–15. DOI: 10.1016/j.ins.2013.12.060. URL: https://doi.org/10.1016/j.ins.2013.12.060.

[161] J. Schmidhuber. "Deep learning in neural networks: An overview". In: *Neural Networks* 61 (Jan. 2015), pp. 85–117. DOI: 10.1016/j.neunet.2014.09.003. URL: https://doi.org/10.1016%2Fj.neunet.2014.09.003.

[162] M. Schuster and K. K. Paliwal. "Bidirectional recurrent neural networks". In: *Neural Computation* 45.11 (1997), pp. 2673–2681. URL: https://maxwell.ict.griffith.edu.au/spl/publications/papers/ieeesp97_schuster.pdf.

[163] M. V. Shcherbakov, A. Brebels, N. L. Shcherbakova, A. P. Tyukov, T. Alex, R. Janovsky, and V. A. Kamaev. "A survey of forecast error measures". In: *World Applied Sciences Journal* 1.24 (Jan. 2013), pp. 171–176. DOI: 10.5829/idosi.wasj.2013.24.itmies.80032. URL: https://www.researchgate.net/publication/281718517_A_survey_of_forecast_error_measures.

[164] A. Shiri, M. Afshar, A. Rahimi-Kian, and B. Maham. "Electricity price forecasting using Support Vector Machines by considering oil and natural gas price impacts". In: *2015 IEEE International Conference on Smart Energy Grid Engineering (SEGE)*. IEEE, Aug. 2015. DOI: 10.1109/sege.2015.7324591. URL: https://doi.org/10.1109/sege.2015.7324591.

[165] N. A. Shrivastava, S. Ch, and B. K. Panigrahi. "Price forecasting using computational intelligence techniques: A comparative analysis". In: *2011 International Conference on Energy, Automation and Signal*. IEEE, Dec. 2011. DOI: 10.1109/iceas.2011.6147107. URL: https://doi.org/10.1109/iceas.2011.6147107.

[166] N. A. Shrivastava, A. Khosravi, and B. K. Panigrahi. "Prediction interval estimation for electricity price and demand using support vector machines". In: *2014 International Joint Conference on Neural Networks (IJCNN)*. IEEE, July 2014. DOI: 10.1109/ijcnn.2014.6889745. URL: https://doi.org/10.1109/ijcnn.2014.6889745.

[167] N. A. Shrivastava, A. Khosravi, and B. K. Panigrahi. "Prediction Interval Estimation of Electricity Prices Using PSO-Tuned Support Vector Machines". In: *IEEE Transactions on Industrial Informatics* 11.2 (Apr. 2015), pp. 322–331. DOI: 10.1109/tii.2015.2389625. URL: https://doi.org/10.1109/tii.2015.2389625.

[168] N. A. Shrivastava and B. K. Panigrahi. "Prediction interval estimations for electricity demands and prices: a multi-objective approach". In: *IET Generation, Transmission & Distribution* 9.5 (Apr. 2015), pp. 494–502. DOI: 10.1049/iet-gtd.2014.0599. URL: https://doi.org/10.1049/iet-gtd.2014.0599.

[169] F. Silva, B. Teixeira, T. Pinto, G. Santos, Z. Vale, and I. Praça. "Generation of realistic scenarios for multi-agent simulation of electricity markets". In: *Energy* 116 (Dec. 2016), pp. 128–139. DOI: 10.1016/j.energy.2016.09.096. URL: http://dx.doi.org/10.1016/j.energy.2016.09.096.

[170] N. K. Singh, M. Tripathy, and A. K. Singh. "A radial basis function neural network approach for multi-hour short term load-price forecasting with type of day parameter". In: *2011 6th International Conference on Industrial and Information Systems*. IEEE, Aug. 2011. DOI: 10.1109/iciinfs.2011.6038087. URL: https://doi.org/10.1109/iciinfs.2011.6038087.

[171] D. Singhal and K. Swarup. "Electricity price forecasting using artificial neural networks". In: *International Journal of Electrical Power & Energy Systems* 33.3 (Mar. 2011), pp. 550–555. DOI: 10.1016/j.ijepes.2010.12.009. URL: https://doi.org/10.1016/j.ijepes.2010.12.009.

[172] R. Skopal. "Short-term hourly price forward curve prediction using neural network and hybrid ARIMA-NN model". In: *2015 International Conference on Information and Digital Technologies*. IEEE, July 2015. DOI: 10.1109/dt.2015.7222993. URL: https://doi.org/10.1109/dt.2015.7222993.

[173] P. Smolensky. "Information processing in dynamical systems: Foundations of harmony theory". In: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations*. Vol. 1. 1986, pp. 194–281. URL: http://smolensky.johnshopkins.edu/docs/Smolensky86PDPch6.

[174] J. Sun and X.-Z. Wang. "An initial comparison on noise resisting between crisp and fuzzy decision trees". In: *2005 International Conference on Machine Learning and Cybernetics*. IEEE, 2005. DOI: 10.1109/icmlc.2005.1527372. URL: http://dx.doi.org/10.1109/ICMLC.2005.1527372.

[175] B. Szkuta, L. Sanabria, and T. Dillon. "Electricity price short-term forecasting using artificial neural networks". In: *IEEE Transactions on Power Systems* 14.3 (1999), pp. 851–857. DOI: 10.1109/59.780895. URL: https://doi.org/10.1109/59.780895.

[176] Z. Tan, J. Zhang, J. Wang, and J. Xu. "Day-ahead electricity price forecasting using wavelet transform combined with ARIMA and GARCH models". In: *Applied Energy* 87.11 (Nov. 2010), pp. 3606–3610. DOI: 10.1016/j.apenergy.2010.05.012. URL: https://doi.org/10.1016/j.apenergy.2010.05.012.

[177] H. Tian, K. Li, and B. Meng. "A novel prediction modeling scheme based on multiple information fusion for day-ahead electricity price". In: *2011 Chinese Control and Decision Conference (CCDC)*. IEEE, May 2011. DOI: 10.1109/ccdc.2011.5968490. URL: https://doi.org/10.1109/ccdc.2011.5968490.

[178] H. Tian and B. Meng. "A new modeling method based on bagging ELM for day-ahead electricity price prediction". In: *2010 IEEE Fifth International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA)*. IEEE, Sept. 2010. DOI: 10.1109/bicta.2010.5645111. URL: https://doi.org/10.1109/bicta.2010.5645111.

[179] H. Tian, B. Meng, and S. Wang. "Day-ahead electricity price prediction based on multiple ELM". In: *2010 Chinese Control and Decision Conference*. IEEE, May 2010. DOI: 10.1109/ccdc.2010.5499079. URL: https://doi.org/10.1109/ccdc.2010.5499079.

[180] H. Tian, M. Zhang, and B. Meng. "Prediction of day-ahead electricity price based on information fusion". In: *2010 International Conference on Computer and Information Application*. IEEE, Dec. 2010. DOI: 10.1109/iccia.2010.6141635. URL: https://doi.org/10.1109/iccia.2010.6141635.

[181] R. Tibshirani. "Regression Shrinkage and Selection via the Lasso". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 58.1 (1996), pp. 267–288. ISSN: 00359246. URL: http://www.jstor.org/stable/2346178.

[182] R. Tibshirani, I. Johnstone, T. Hastie, and B. Efron. "Least angle regression". In: *The Annals of Statistics* 32.2 (Apr. 2004), pp. 407–499. DOI: 10.1214/009053604000000067. URL: https://doi.org/10.1214/009053604000000067.

[183] T. Tieleman and G. Hinton. *Lecture 6.5—RmsProp: Divide the gradient by a running average of its recent magnitude*. COURSERA: Neural Networks for Machine Learning. 2012. URL: http://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf.

[184] C. Tofallis. "A better measure of relative prediction accuracy for model selection and model estimation". In: *Journal of the Operational Research Society* 66.8 (Nov. 2014), pp. 1352–1362. DOI: 10.1057/jors.2014.103. URL: https://doi.org/10.1057/jors.2014.103.

[185] S. S. Torbaghan, A. Motamedi, H. Zareipour, and L. A. Tuan. "Medium-term electricity price forecasting". In: *2012 North American Power Symposium (NAPS)*. IEEE, Sept. 2012. DOI: 10.1109/naps.2012.6336424. URL: https://doi.org/10.1109/naps.2012.6336424.

[186] M. A. Torkamani and D. Lowd. "On Robustness and Regularization of Structural Support Vector Machines". In: *Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32*. ICML'14. Beijing, China: JMLR.org, 2014, pp. II-577–II-585. URL: http://proceedings.mlr.press/v32/torkamani14.pdf.

[187] T. B. Trafalis and R. C. Gilbert. "Robust classification and regression using support vector machines". In: *European Journal of Operational Research* 173.3 (Sept. 2006), pp. 893–909. DOI: 10.1016/j.ejor.2005.07.024. URL: https://doi.org/10.1016/j.ejor.2005.07.024.

[188] F. van Veen. *The neural network zoo*. [Online; accessed 2017-02-28]. 2016. URL: http://www.asimovinstitute.org/neural-network-zoo/.

[189] Y. Vardanyan and M. R. Hesamzadeh. "Modeling regime switching in day-ahead market prices using Markov model". In: *2016 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*. IEEE, Oct. 2016. DOI: 10.1109/isgteurope.2016.7856316. URL: https://doi.org/10.1109/isgteurope.2016.7856316.

[190] J. Vert, K. Tsuda, and B. Schölkopf. "A Primer on Kernel Methods". In: *Kernel Methods in Computational Biology*. Cambridge, MA, USA: MIT Press, 2004, pp. 35–70.

[191] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol. "Extracting and Composing Robust Features with Denoising Autoencoders". In: *Proceedings of the 25th International Conference on Machine Learning*. ICML '08. Helsinki, Finland: ACM, 2008, pp. 1096–1103. ISBN: 978-1-60558-205-4. DOI: 10.1145/1390156.1390294. URL: http://doi.acm.org/10.1145/1390156.1390294.

[192] C. Wan, M. Niu, Y. Song, and Z. Xu. "Pareto Optimal Prediction Intervals of Electricity Price". In: *IEEE Transactions on Power Systems* 32.1 (Jan. 2017), pp. 817–819. DOI: 10.1109/tpwrs.2016.2550867. URL: https://doi.org/10.1109/tpwrs.2016.2550867.

[193] C. Wan, Z. Xu, Y. Wang, Z. Y. Dong, and K. P. Wong. "A Hybrid Approach for Probabilistic Forecasting of Electricity Price". In: *IEEE Transactions on Smart Grid* 5.1 (Jan. 2014), pp. 463–470. DOI: 10.1109/tsg.2013.2274465. URL: https://doi.org/10.1109/tsg.2013.2274465.

[194] A. Wang and B. Ramsay. "A neural network based estimator for electricity spot-pricing with particular reference to weekend and public holidays". In: *Neurocomputing* 23.1-3 (Dec. 1998), pp. 47–57. DOI: 10.1016/s0925-2312(98)00079-4. URL: https://doi.org/10.1016/s0925-2312(98)00079-4.

[195] Q. Wang, Z. Dong, X. Li, J. Zhao, and K. P. Wong. "An Effective Approach to Predicting Electricity Market Price Spikes". In: *2007 IEEE Power Engineering Society General Meeting*. Institute of Electrical & Electronics Engineers (IEEE), June 2007. DOI: 10.1109/pes.2007.385852. URL: http://dx.doi.org/10.1109/PES.2007.385852.

[196] P. J. Werbos. "Applications of advances in nonlinear sensitivity analysis". In: *System Modeling and Optimization*. Springer-Verlag, 1981, pp. 762–770. DOI: 10.1007/bfb0006203. URL: https://doi.org/10.1007%2Fbfb0006203.

[197] R. Weron. "Electricity price forecasting: A review of the state-of-the-art with a look into the future". In: *International Journal of Forecasting* 30.4 (Oct. 2014), pp. 1030–1081. DOI: 10.1016/j.ijforecast.2014.08.008. URL: http://dx.doi.org/10.1016/j.ijforecast.2014.08.008.

[198] R. Weron and A. Misiorek. "Forecasting Spot Electricity Prices With Time Series Models". In: *Proceedings of international conference on the European electricity market EEM. Lodz, Poland; 10–12 May 2005*. 2005. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.126.8406.

[199] J. M. Wooldridge. *Introductory Econometrics - A modern approach*. 4th ed. Cengage Learning Emea, 2008.

[200] H. C. Wu, S. C. Chan, K. M. Tsui, and Y. Hou. "A New Recursive Dynamic Factor Analysis for Point and Interval Forecast of Electricity Price". In: *IEEE Transactions on Power Systems* 28.3 (Aug. 2013), pp. 2352–2365. DOI: 10.1109/tpwrs.2012.2232314. URL: https://doi.org/10.1109/tpwrs.2012.2232314.

[201] L. Wu and M. Shahidehpour. "A Hybrid Model for Day-Ahead Price Forecasting". In: *IEEE Transactions on Power Systems* 25.3 (Aug. 2010), pp. 1519–1530. DOI: 10.1109/tpwrs.2009.2039948. URL: https://doi.org/10.1109/tpwrs.2009.2039948.

[202] W. Wu, J. Zhou, L. Mo, and C. Zhu. "Forecasting Electricity Market Price Spikes Based on Bayesian Expert with Support Vector Machines". In: *Advanced Data Mining and Applications*. Springer Berlin Heidelberg, 2006, pp. 205–212. DOI: 10.1007/11811305_23. URL: https://doi.org/10.1007/11811305_23.

[203] W. Wu, J.-Z. Zhou, J. Yu, C.-J. Zhu, and J.-J. Yang. "Prediction of spot market prices of electricity using chaotic time series". In: *Proceedings of 2004 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.04EX826)*. IEEE, 2004. DOI: 10.1109/icmlc.2004.1382311. URL: https://doi.org/10.1109/icmlc.2004.1382311.

[204] M. Xie, C. Sandels, K. Zhu, and L. Nordstrom. "A seasonal ARIMA model with exogenous variables for elspot electricity prices in Sweden". In: *2013 10th International Conference on the European Energy Market (EEM)*. IEEE, May 2013. DOI: 10.1109/eem.2013.6607293. URL: https://doi.org/10.1109/eem.2013.6607293.

[205] H. Xu and T. Niimura. "Short-term electricity price modeling and forecasting using wavelets and multivariate time series". In: *IEEE PES Power Systems Conference and Exposition*. IEEE, 2004. DOI: 10.1109/psce.2004.1397570. URL: https://doi.org/10.1109/psce.2004.1397570.

[206] H. Xu, S. Mannor, and C. Caramanis. "Robustness, Risk, and Regularization in Support Vector Machines". In: *CoRR* abs/0803.3490 (2008). URL: http://arxiv.org/abs/0803.3490.

[207] H. Y. Yamin, S. Shahidehpour, and Z. LI. "Adaptive short-term electricity price forecasting using artificial neural networks in the restructured power markets". In: *International Journal of Electrical Power & Energy Systems* 26.8 (Oct. 2004), pp. 571–581. DOI: 10.1016/j.ijepes.2004.04.005. URL: https://doi.org/10.1016/j.ijepes.2004.04.005.

[208] J. Yan, C. Tian, Y. Wang, and J. Huang. "Online incremental regression for electricity price prediction". In: *Proceedings of 2012 IEEE International Conference on Service Operations and Logistics, and Informatics*. IEEE, July 2012. DOI: 10.1109/soli.2012.6273500. URL: https://doi.org/10.1109/soli.2012.6273500.

[209] H. Yang and X. Duan. "Chaotic characteristics of electricity price and its forecasting model". In: *CCECE 2003 - Canadian Conference on Electrical and Computer Engineering. Toward a Caring and Humane Technology (Cat. No.03CH37436)*. IEEE, 2003. DOI: 10.1109/ccece.2003.1226481. URL: https://doi.org/10.1109/ccece.2003.1226481.

[210] S. J. Yao, Y. H. Song, and X. L. Z. Zhang. "Prediction of System Marginal Prices by Wavelet Transform and Neural Networks". In: *Electric Machines & Power Systems* 28.10 (Oct. 2000), pp. 983–993. DOI: 10.1080/07313560050129855. URL: https://doi.org/10.1080/07313560050129855.

[211] W. Zaremba and I. Sutskever. *Reinforcement Learning Neural Turing Machines - Revised*. 2015. eprint: arXiv:1505.00521. URL: https://arxiv.org/abs/1505.00521.

[212] M. D. Zeiler. "ADADELTA: An Adaptive Learning Rate Method". In: *CoRR* abs/1212.5701 (2012). URL: http://arxiv.org/abs/1212.5701.

[213] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus. "Deconvolutional networks". In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Institute of Electrical and Electronics Engineers (IEEE), June 2010. DOI: 10.1109/cvpr.2010.5539957. URL: https://doi.org/10.1109%2Fcvpr.2010.5539957.

[214] J. Zhang, J. Han, R. Wang, and G. Hou. "Day-ahead electricity price forecasting based on rolling time series and least square-support vector machine model". In: *2011 Chinese Control and Decision Conference (CCDC)*. IEEE, May 2011. DOI: 10.1109/ccdc.2011.5968342. URL: https://doi.org/10.1109/ccdc.2011.5968342.

[215] J. Zhang, F. Wang, R. Wang, and G. Hou. "A study on the application of BPNN based on Minimum Error Entropy in electricity price forecasting". In: *2011 6th IEEE Conference on Industrial Electronics and Applications*. IEEE, June 2011. DOI: 10.1109/iciea.2011.5975719. URL: https://doi.org/10.1109/iciea.2011.5975719.

[216] L. Zhang and P. Luh. "Neural Network-Based Market Clearing Price Prediction and Confidence Interval Estimation With an Improved Extended Kalman Filter Method". In: *IEEE Transactions on Power Systems* 20.1 (Feb. 2005), pp. 59–66. DOI: 10.1109/tpwrs.2004.840416. URL: https://doi.org/10.1109/tpwrs.2004.840416.

[217] L. Zhang, P. Luh, and K. Kasiviswanathan. "Energy clearing price prediction and confidence interval estimation with cascaded neural networks". In: *IEEE Transactions on Power Systems* 18.1 (Feb. 2003), pp. 99–105. DOI: 10.1109/tpwrs.2002.807062. URL: https://doi.org/10.1109/tpwrs.2002.807062.

[218] Y. Zhang, B. Zhang, F. Coenenz, and W. Lu. "Highly reliable breast cancer diagnosis with cascaded ensemble classifiers". In: *The 2012 International Joint Conference on Neural Networks (IJCNN)*. IEEE, June 2012. DOI: 10.1109/ijcnn.2012.6252547. URL: http://dx.doi.org/10.1109/IJCNN.2012.6252547.

[219] J. H. Zhao, Z. Y. Dong, X. Li, and K. P. Wong. "A Framework for Electricity Price Spike Analysis With Advanced Data Mining Methods". In: *IEEE Transactions on Power Systems* 22.1 (Feb. 2007), pp. 376–385. DOI: 10.1109/tpwrs.2006.889139. URL: http://dx.doi.org/10.1109/TPWRS.2006.889139.

[220] M. Zhou, Z. Yan, Y. Ni, G. Li, and Y. Nie. "Electricity price forecasting with confidence-interval estimation through an extended ARIMA approach". In: *IEE Proceedings - Generation, Transmission and Distribution* 153.2 (2006), p. 187. DOI: 10.1049/ip-gtd:20045131. URL: https://doi.org/10.1049/ip-gtd:20045131.

[221] M. Zhou, Z. Yan, Y. Ni, and G. Li. "An ARIMA approach to forecasting electricity price with accuracy improvement by predicted errors". In: *IEEE Power Engineering Society General Meeting, 2004*. IEEE, 2004. DOI: 10.1109/pes.2004.1372791. URL: https://doi.org/10.1109/pes.2004.1372791.

[222] H. Zou and T. Hastie. "Regularization and Variable Selection via the Elastic Net". In: *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 67.2 (2005), pp. 301–320. ISSN: 13697412, 14679868. URL: http://www.jstor.org/stable/3647580.

Appendices

Appendix A

Violin plots

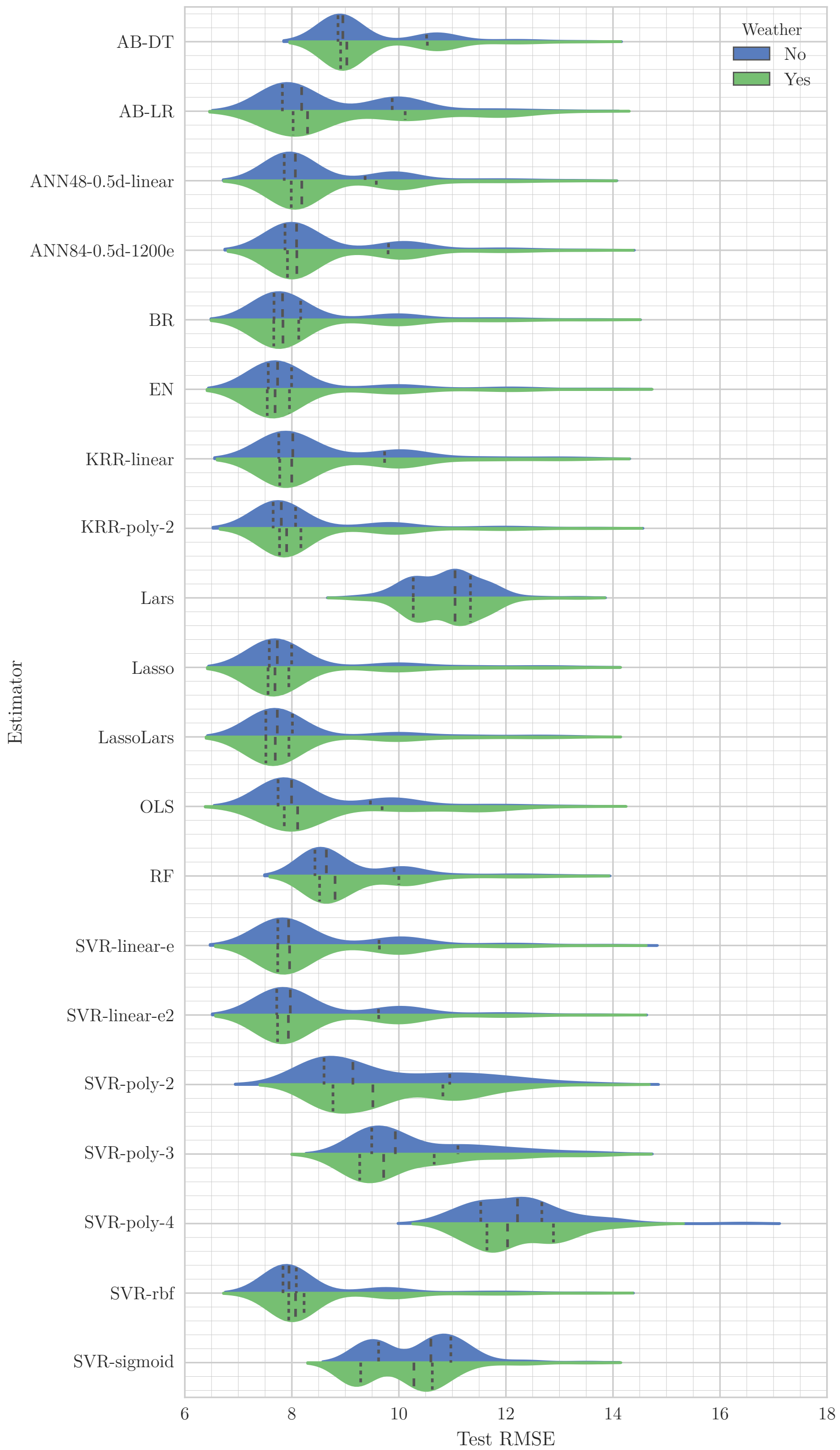


Figure A.1: Distribution of RMSE broken down by individual estimator types and the inclusion of weather data.

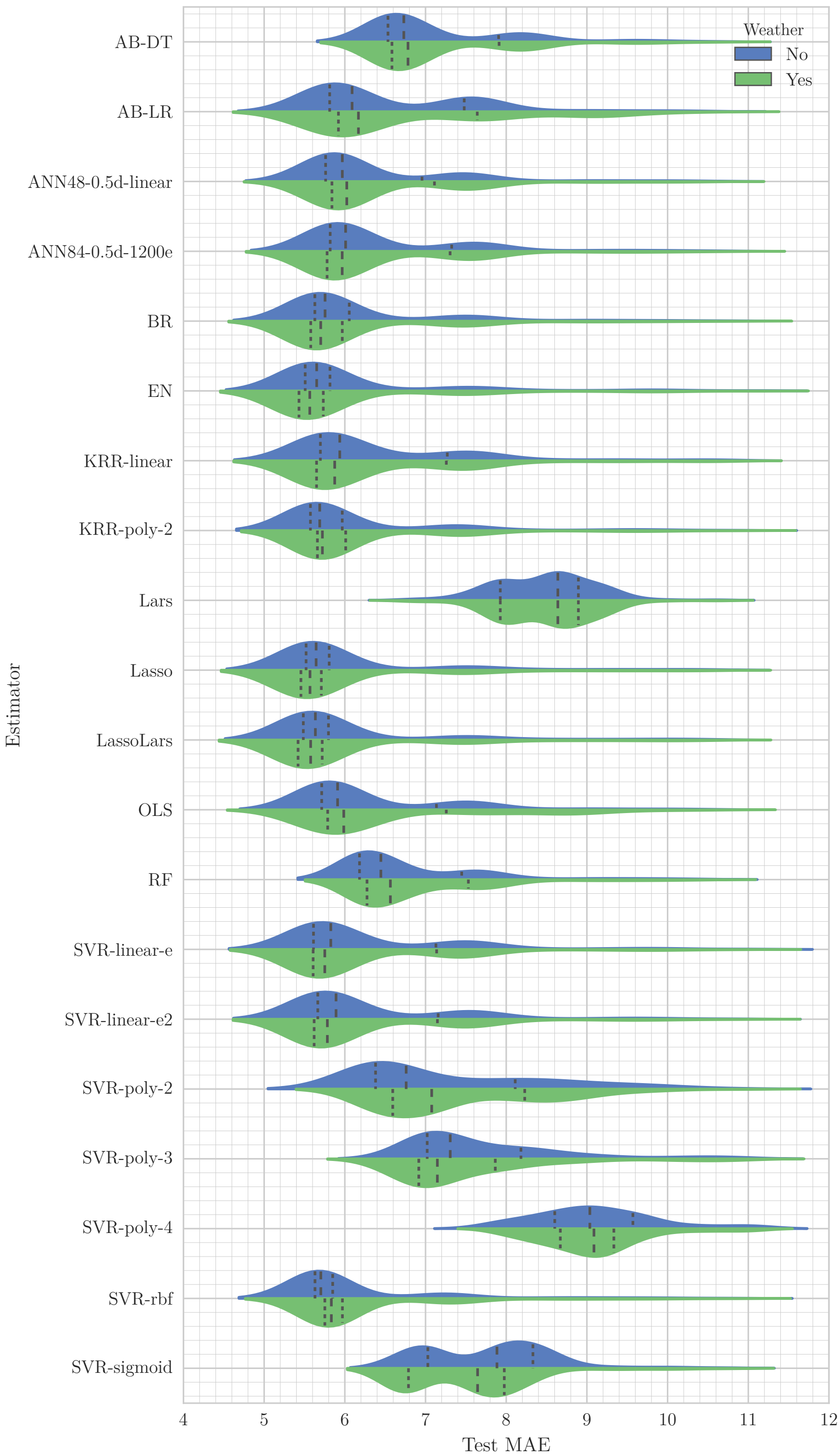


Figure A.2: Distribution of MAE broken down by individual estimator types and the inclusion of weather data.

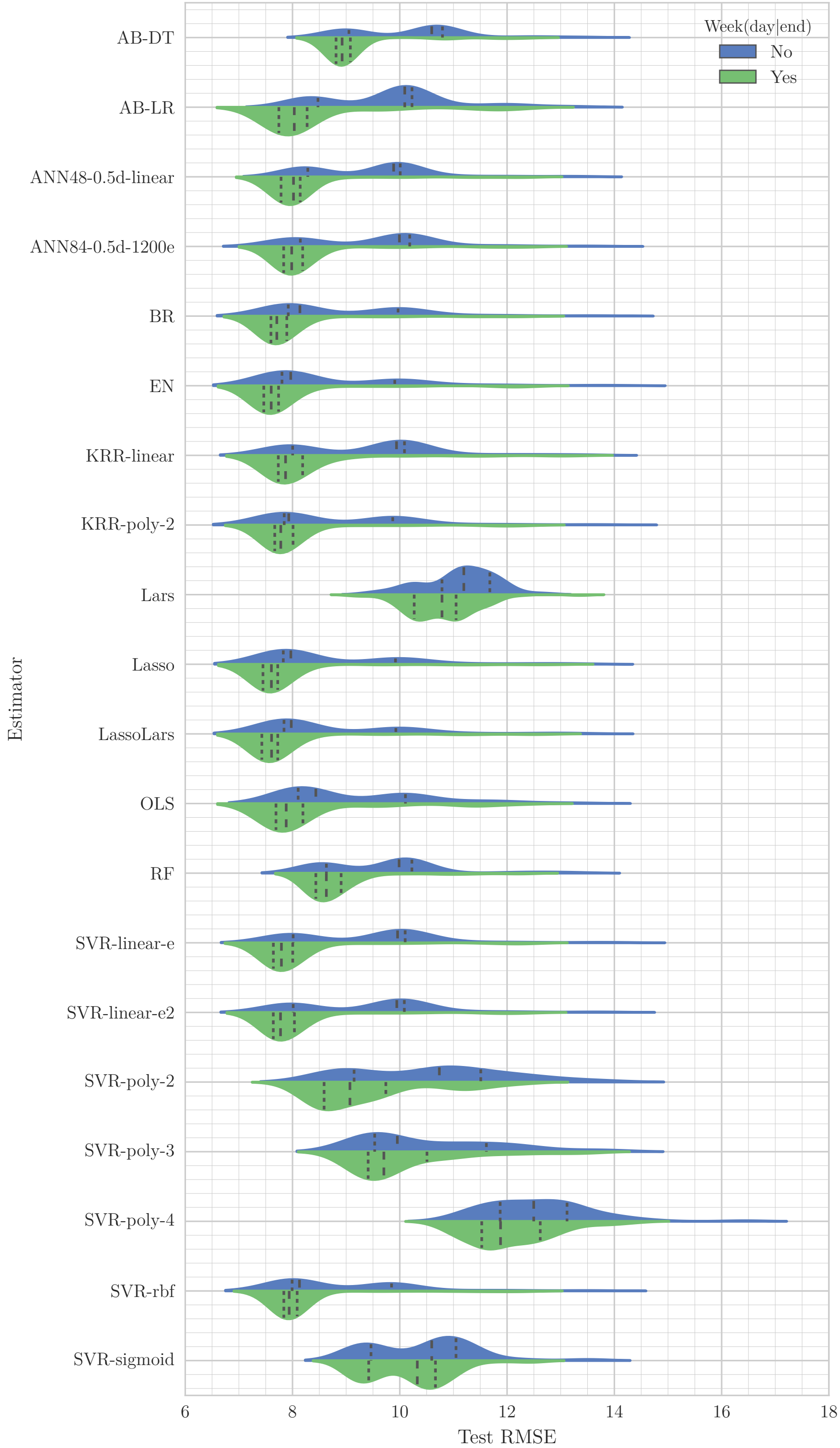


Figure A.3: Distribution of RMSE broken down by individual estimator types and the inclusion of Weekday or Weekend dummy variable. The Weekend variables consists of single dummy variable with value 1 for either Saturday or Sunday. The Weekday variables consists of dummy variables for days from Tuesday to Sunday, Monday is the base day.

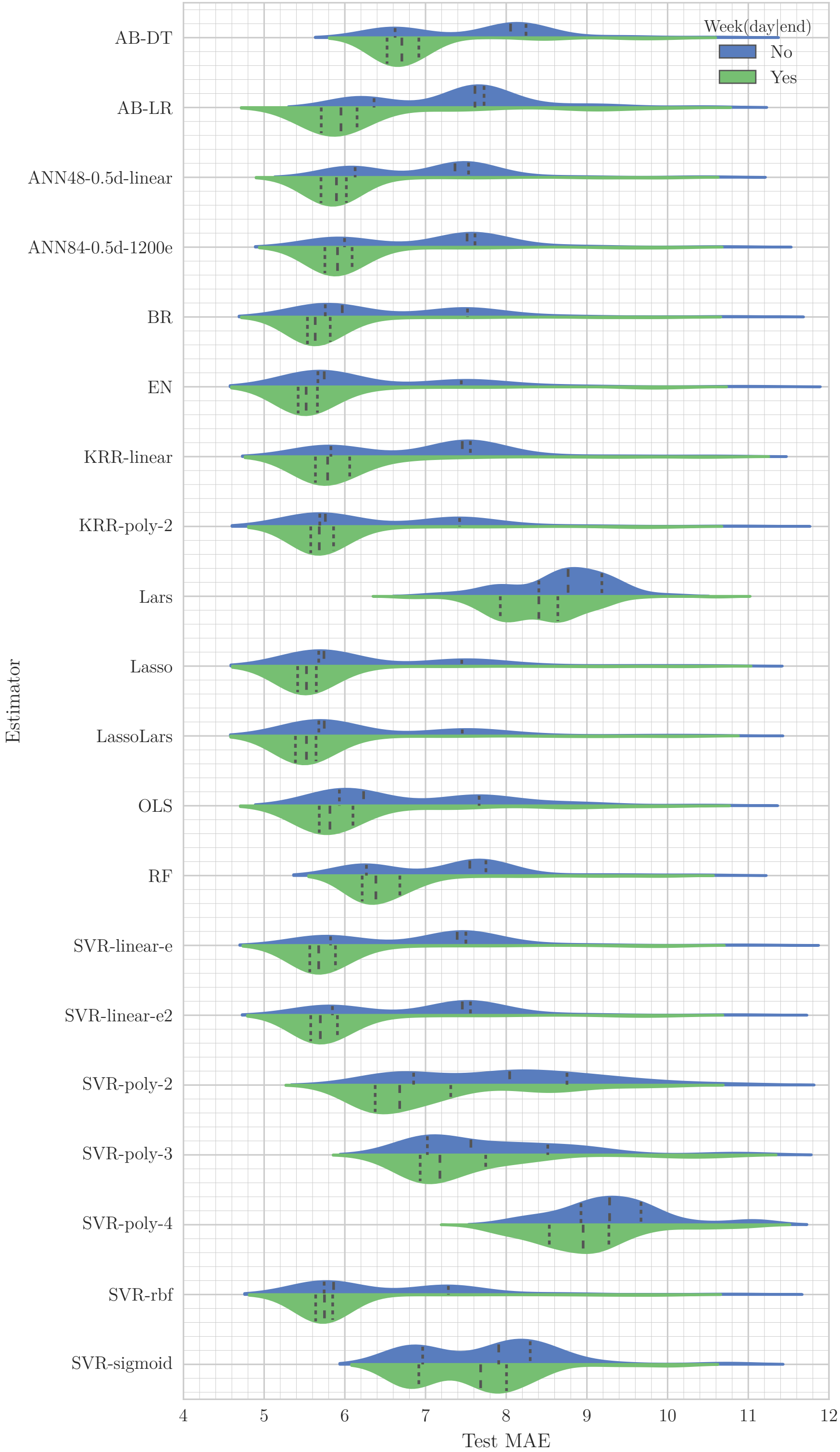


Figure A.4: Distribution of MAE broken down by individual estimator types and the inclusion of Weekday or Weekend dummy variable. The Weekend variables consists of single dummy variable with value 1 for either Saturday or Sunday. The Weekday variables consists of dummy variables for days from Tuesday to Sunday, Monday is the base day.

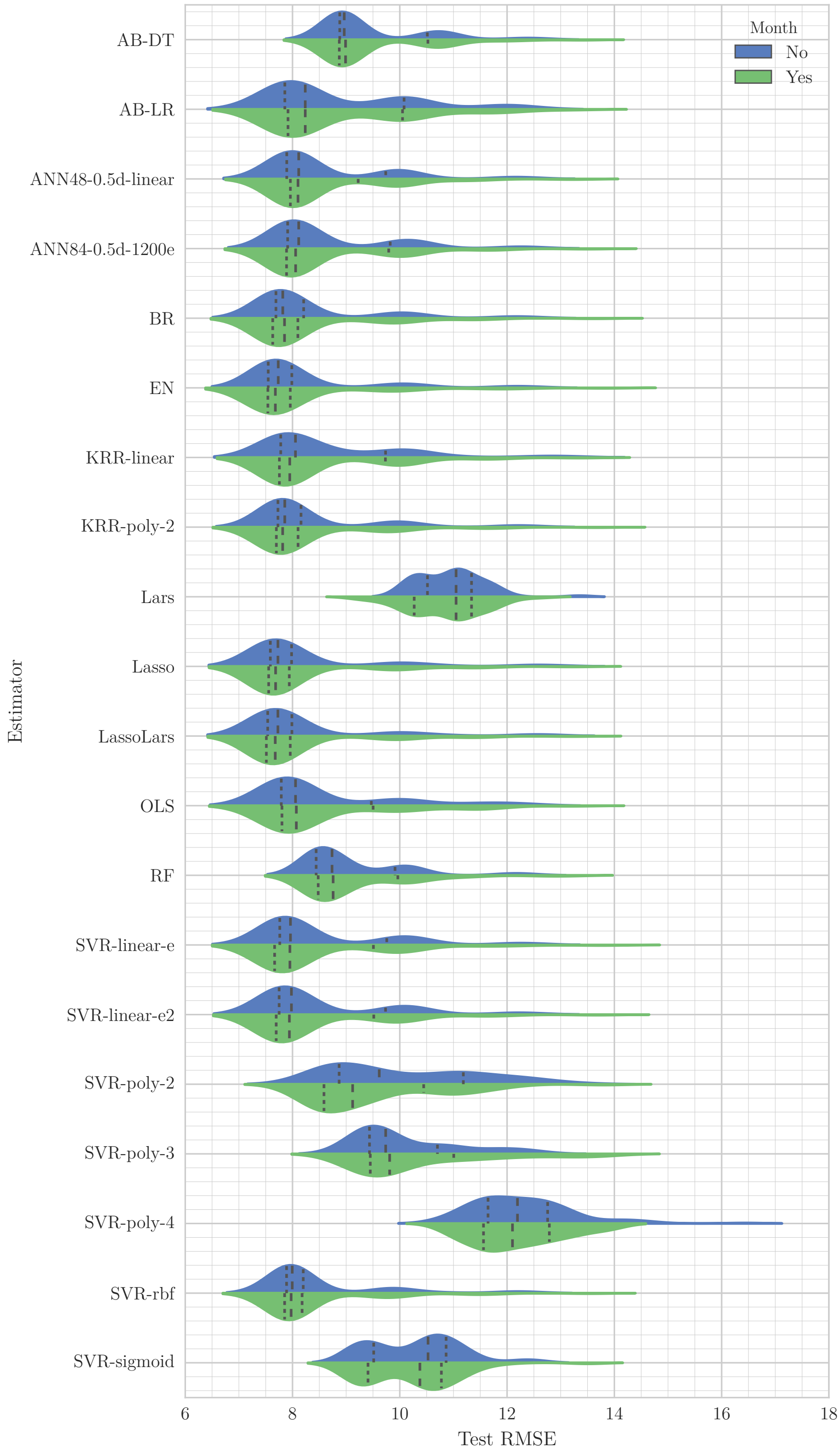


Figure A.5: Distribution of RMSE broken down by individual estimator types and the inclusion of Monthly dummy variables. The Monthly dummy consists of dummy variables for month from February to December, January is the base month.

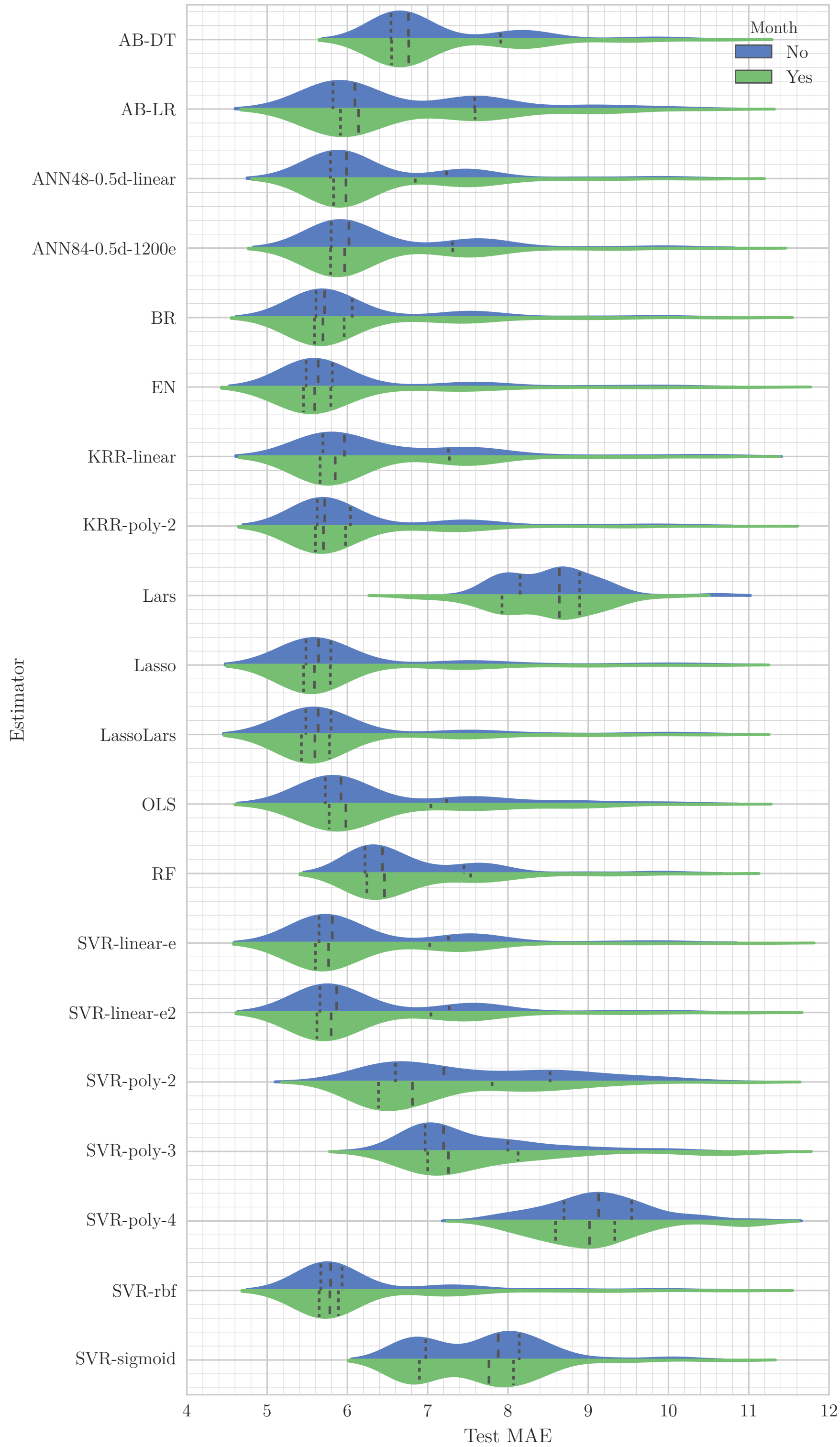


Figure A.6: Distribution of MAE broken down by individual estimator types and the inclusion of Monthly Dummy variables. The Monthly dummy consists of dummy variables for month from February to December, January is the base month.

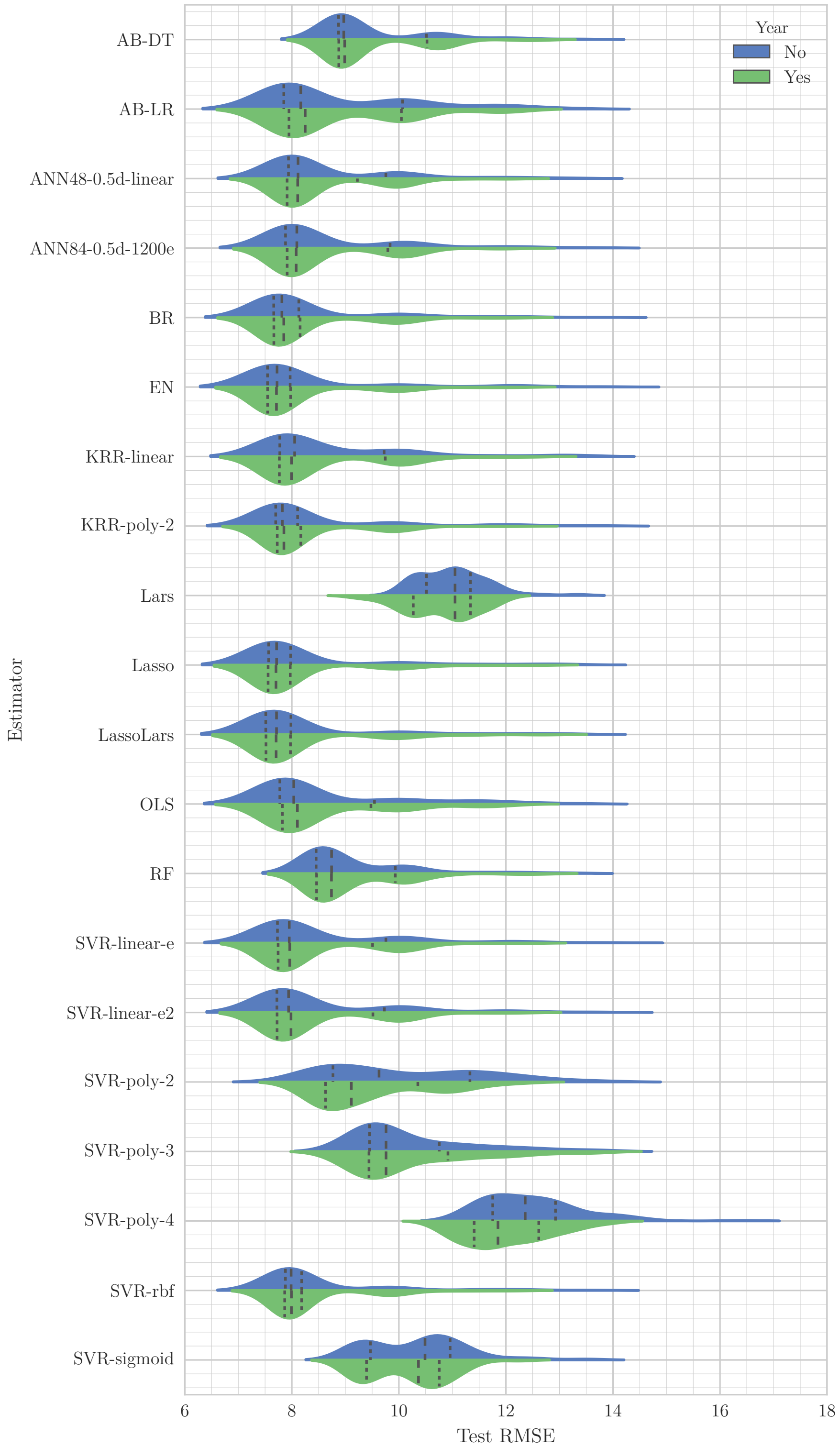


Figure A.7: Distribution of RMSE broken down by individual estimator types and the inclusion of Yearly dummy variables. The Yearly dummy consists of dummy variables for years from 2011 to 2016, 2010 is the base year.

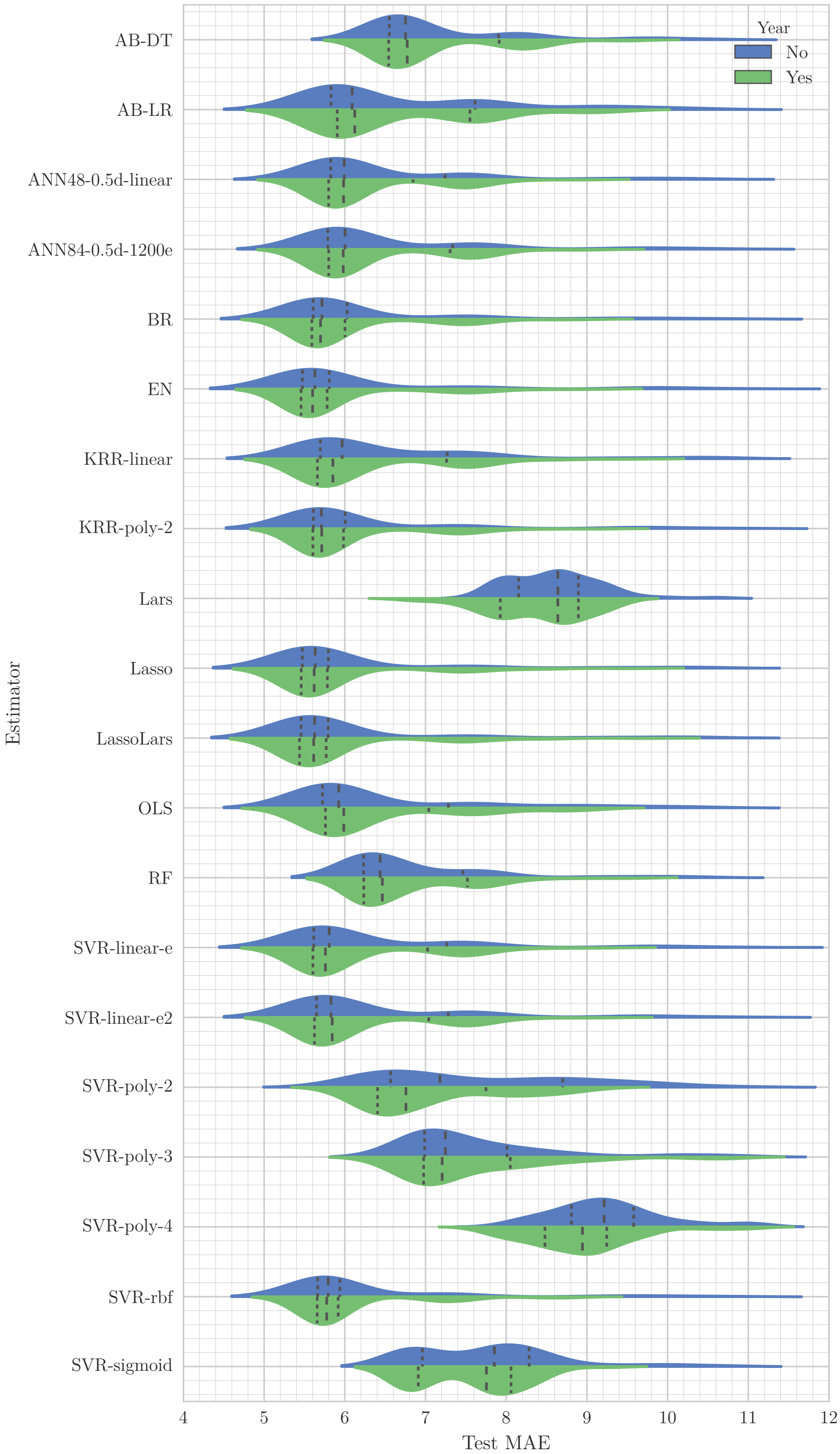


Figure A.8: Distribution of MAE broken down by individual estimator types and the inclusion of Yearly Dummy variables. The Yearly dummy consists of dummy variables for years from 2011 to 2016, 2010 is the base year.

Appendix B

Training and testing errors

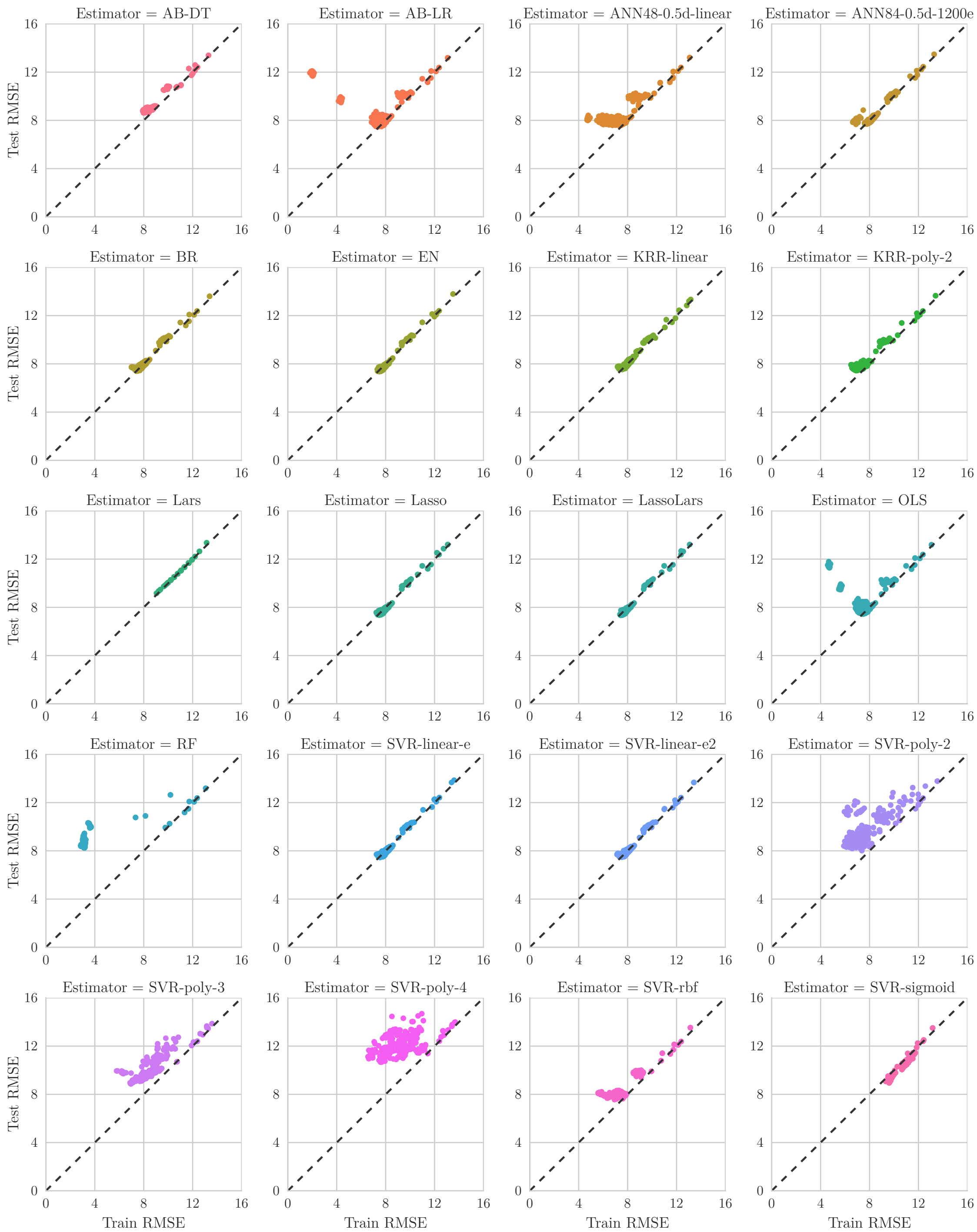


Figure B.1: The training and testing RMSE by individual estimators, each point represent a single estimator.

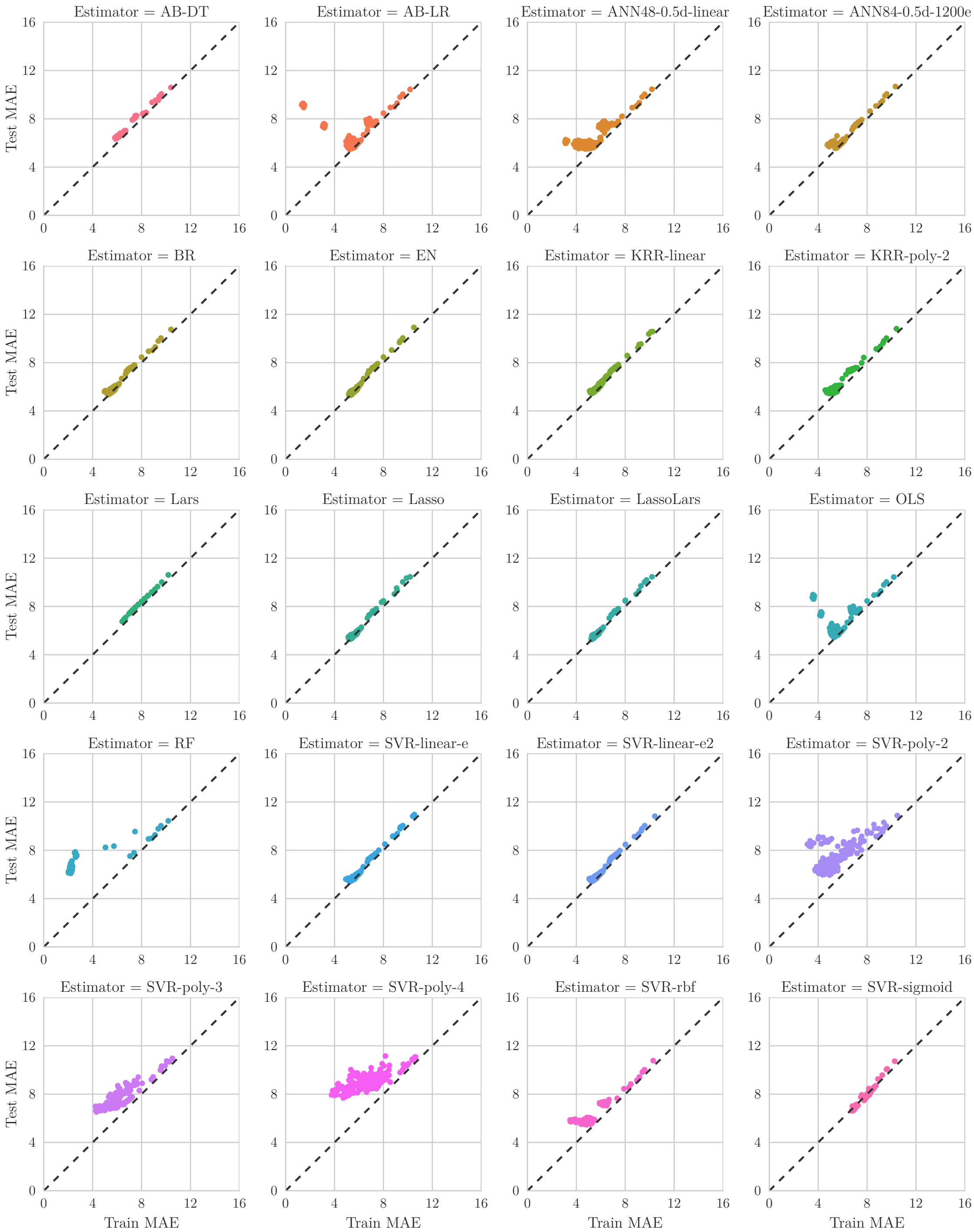
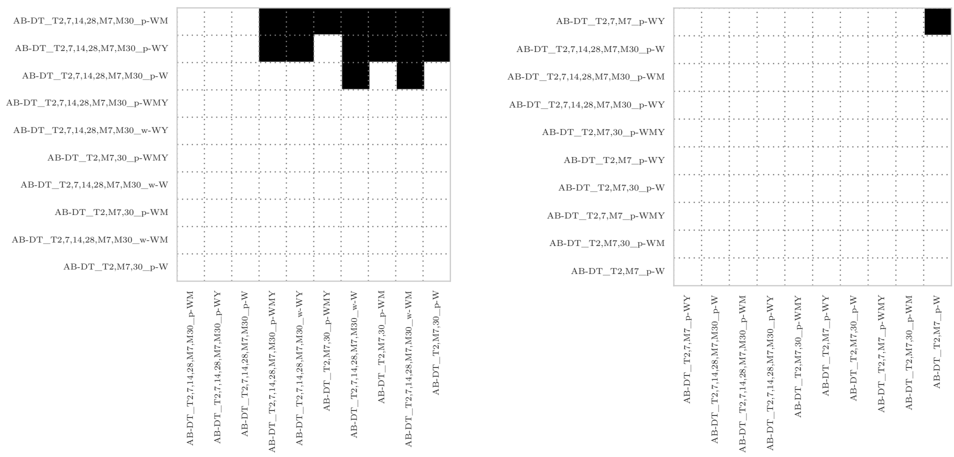


Figure B.2: The training and testing MAE by individual estimators, each point represent a single estimator.

Appendix C

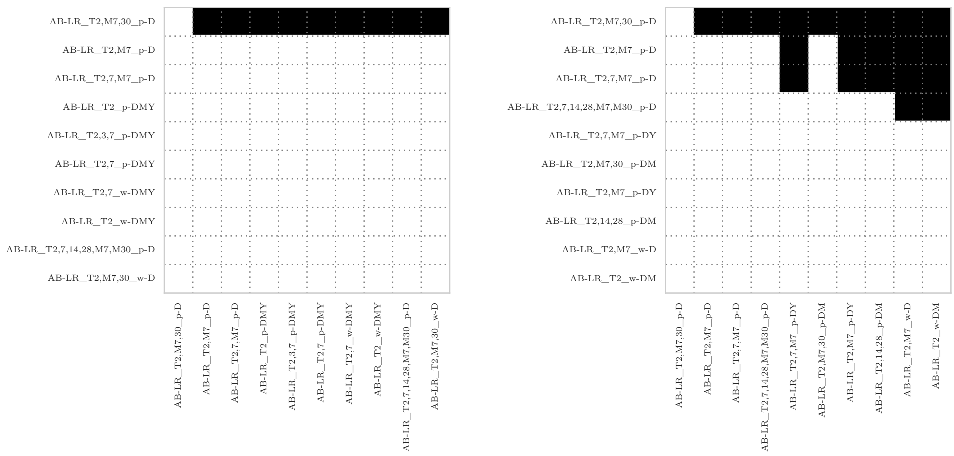
10 best estimators

C.1 DM tests



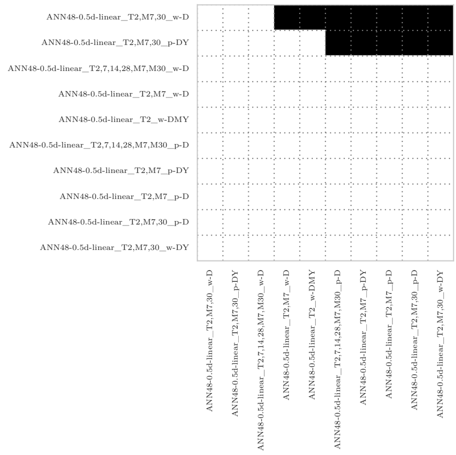
MAE, DM test with $p = 1$ RMSE, DM test with $p = 2$

a `AB-DT`

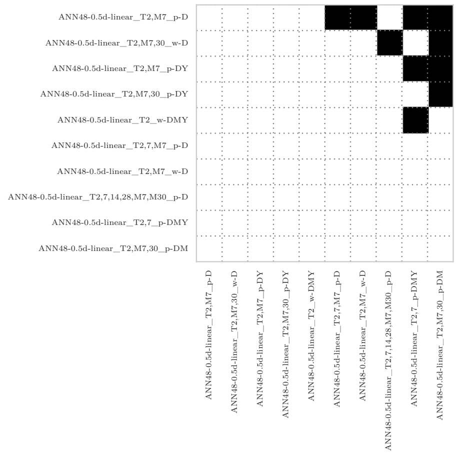


MAE, DM test with $p = 1$ RMSE, DM test with $p = 2$

b `AB-LR`

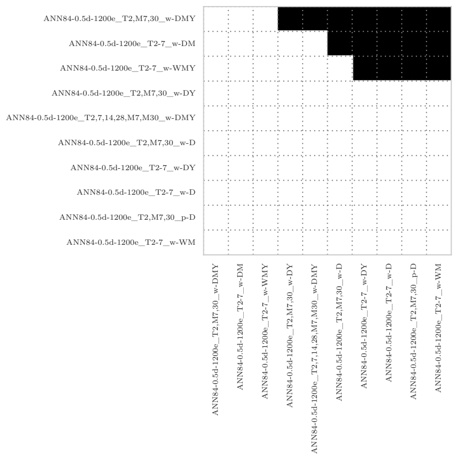


MAE, DM test with $p = 1$

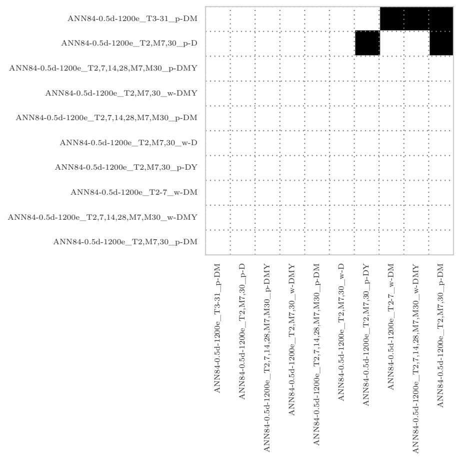


RMSE, DM test with $p = 2$

c `ANN48-0.5d-linear`

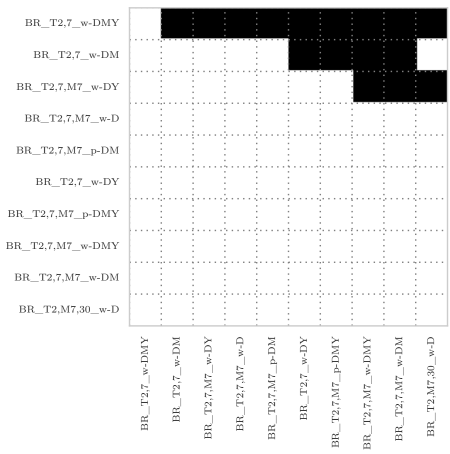


MAE, DM test with $p = 1$

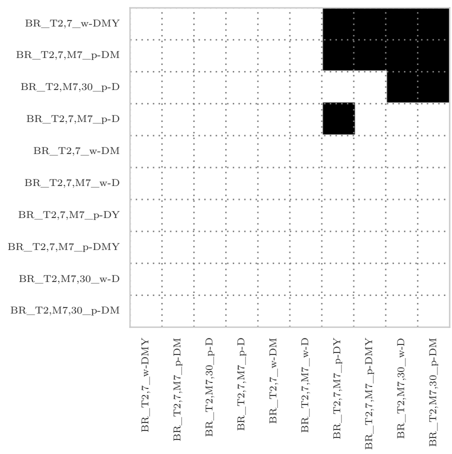


RMSE, DM test with $p = 2$

d `ANN84-0.5d-1200e`

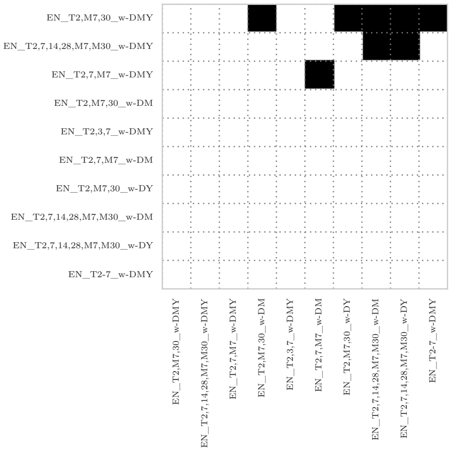


MAE, DM test with $p = 1$

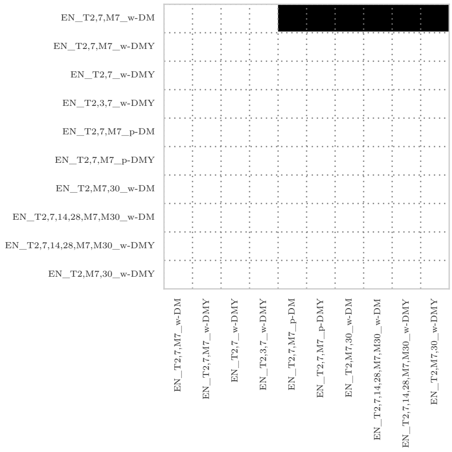


RMSE, DM test with $p = 2$

e `BR`

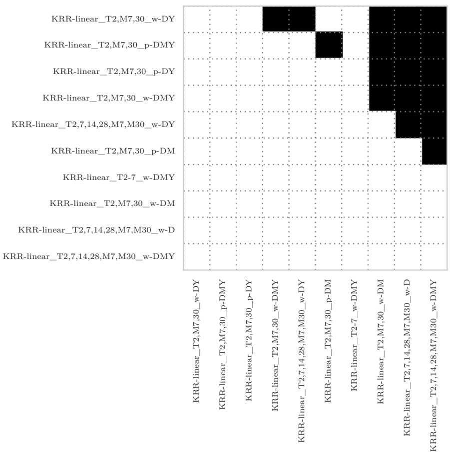


MAE, DM test with $p = 1$

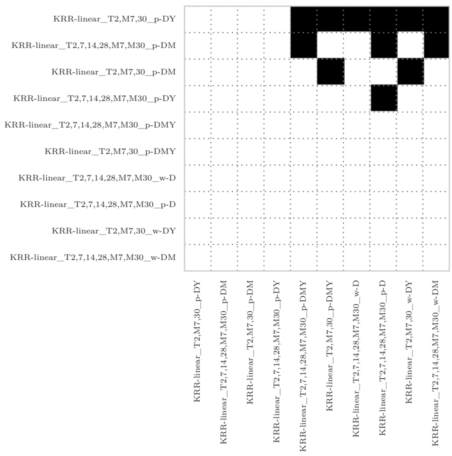


RMSE, DM test with $p = 2$

f `EN`

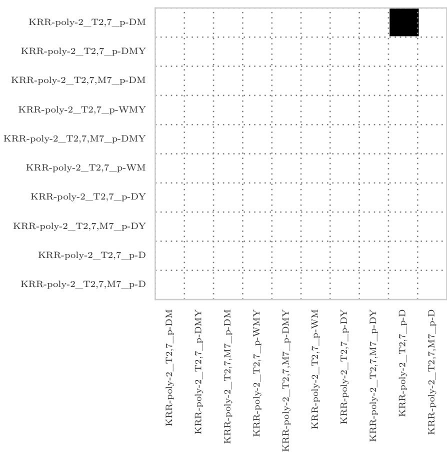


MAE, DM test with $p = 1$

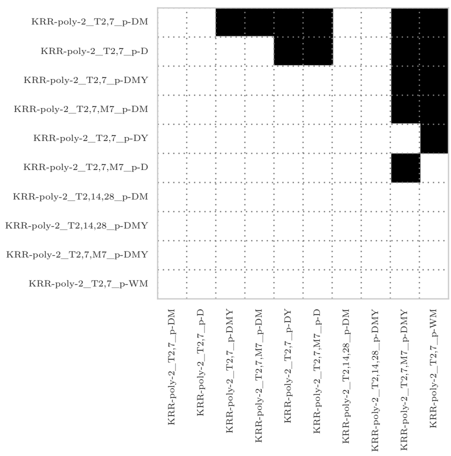


RMSE, DM test with $p = 2$

g `KRR-linear`

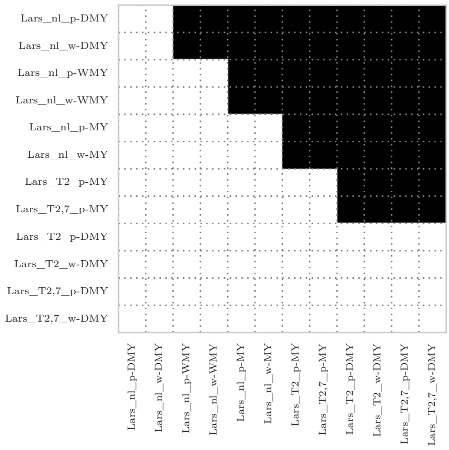


MAE, DM test with $p = 1$

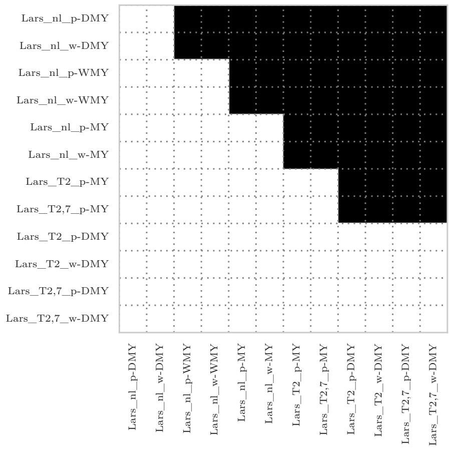


RMSE, DM test with $p = 2$

h `KRR-poly-2`

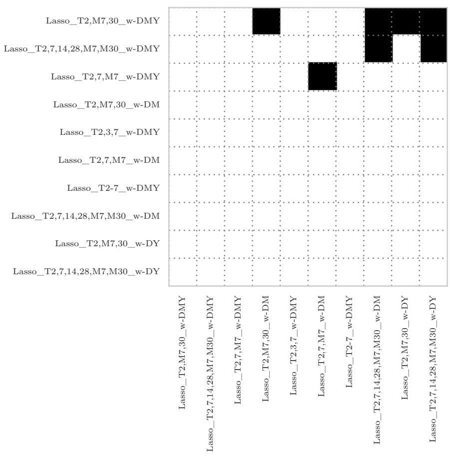


MAE, DM test with $p = 1$

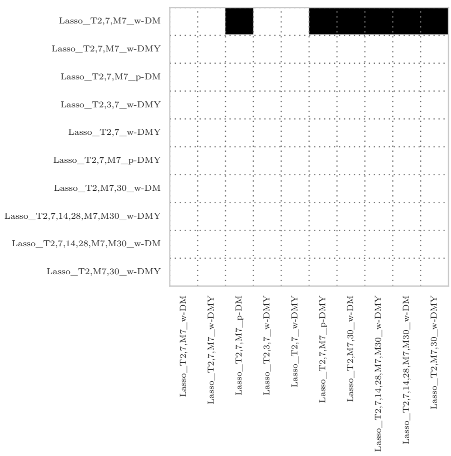


RMSE, DM test with $p = 2$

i Lars

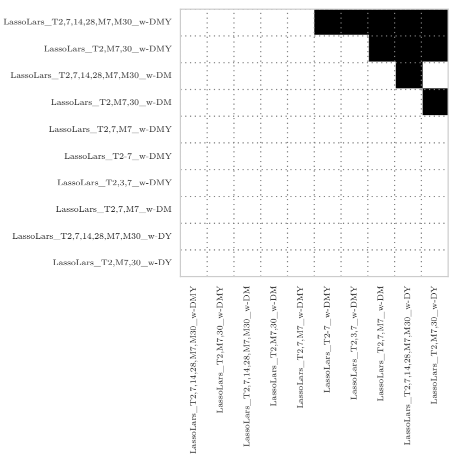


MAE, DM test with $p = 1$

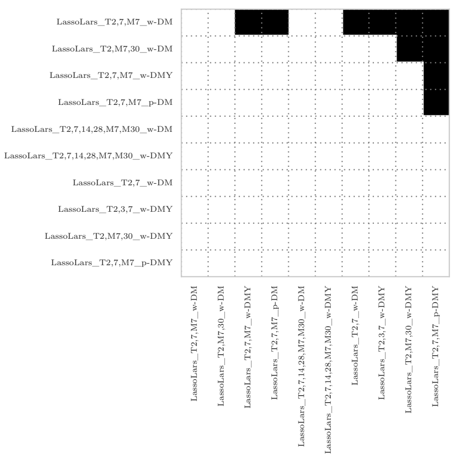


RMSE, DM test with $p = 2$

j Lasso

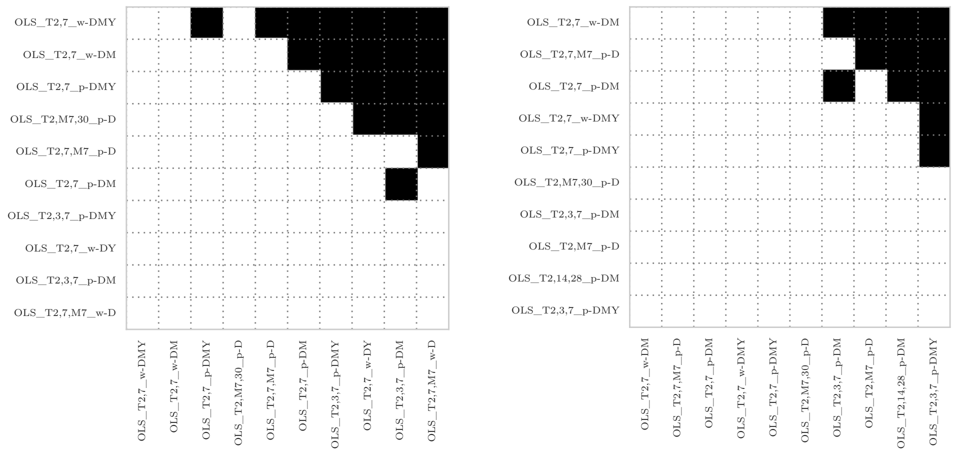


MAE, DM test with $p = 1$



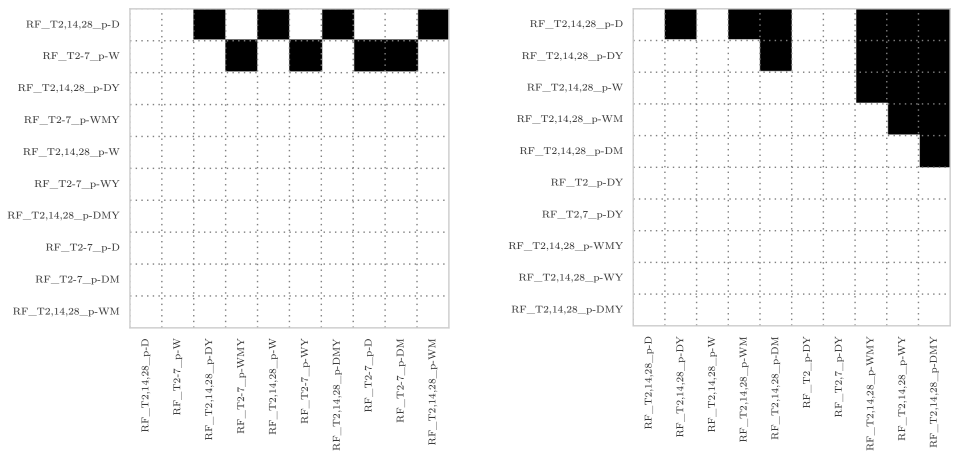
RMSE, DM test with $p = 2$

k LassoLars



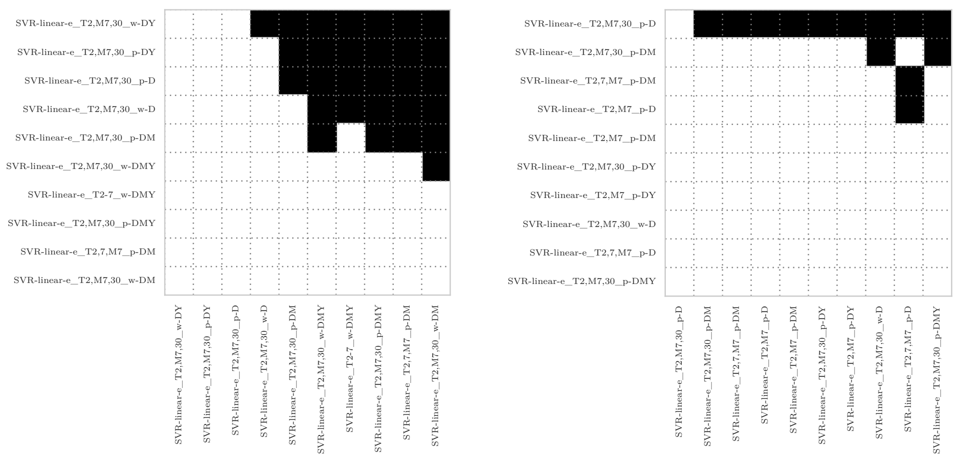
MAE, DM test with $p = 1$ RMSE, DM test with $p = 2$

l OLS



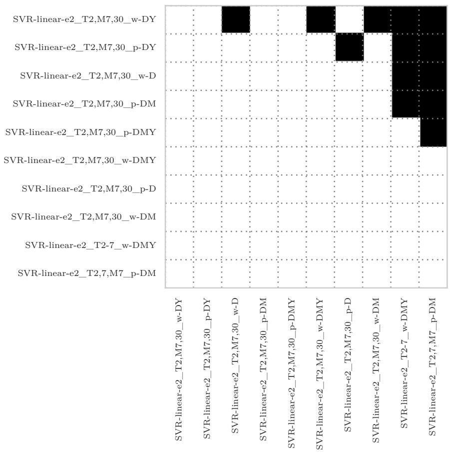
MAE, DM test with $p = 1$ RMSE, DM test with $p = 2$

m RF

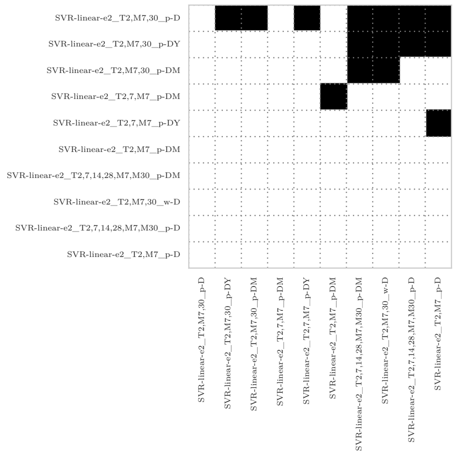


MAE, DM test with $p = 1$ RMSE, DM test with $p = 2$

n SVR-linear-e

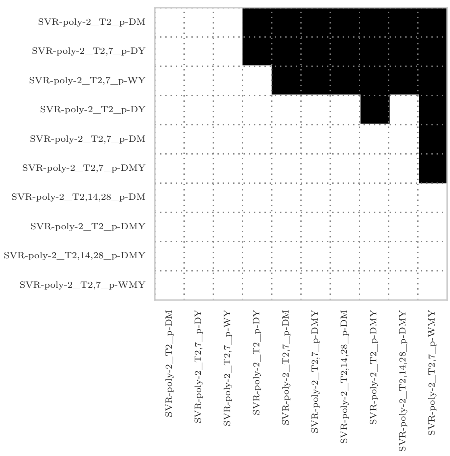


MAE, DM test with $p = 1$

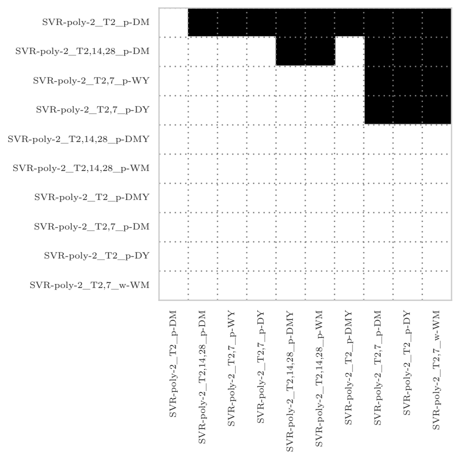


RMSE, DM test with $p = 2$

o `SVR-linear-e2`

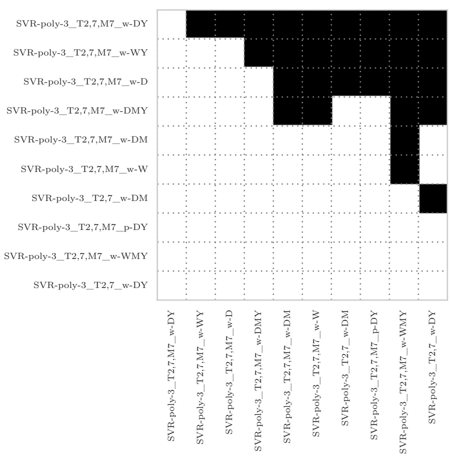


MAE, DM test with $p = 1$

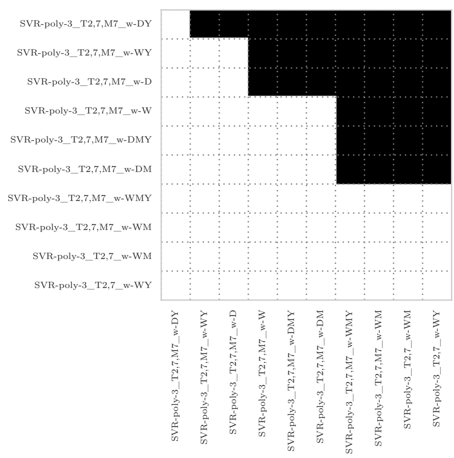


RMSE, DM test with $p = 2$

p `SVR-poly-2`

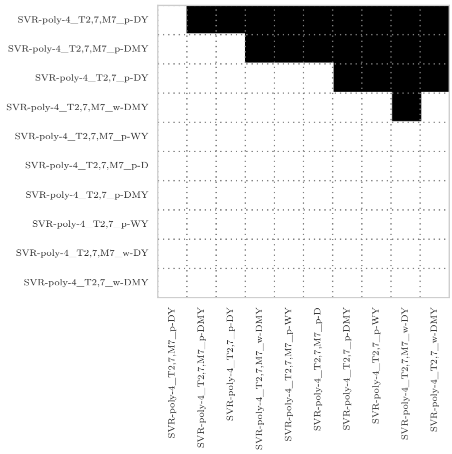


MAE, DM test with $p = 1$

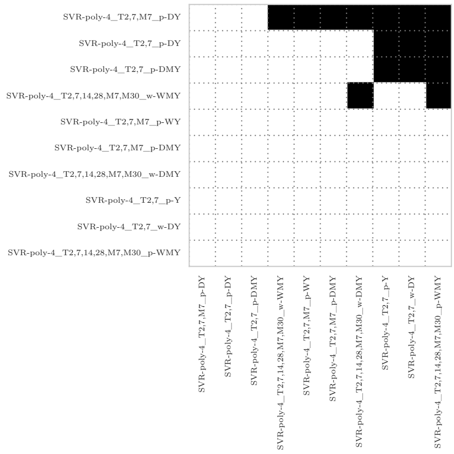


RMSE, DM test with $p = 2$

q `SVR-poly-3`

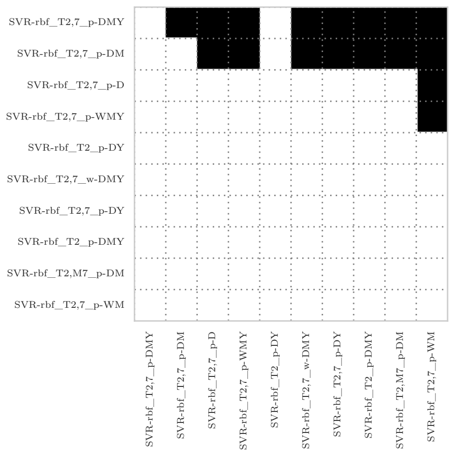


MAE, DM test with $p = 1$

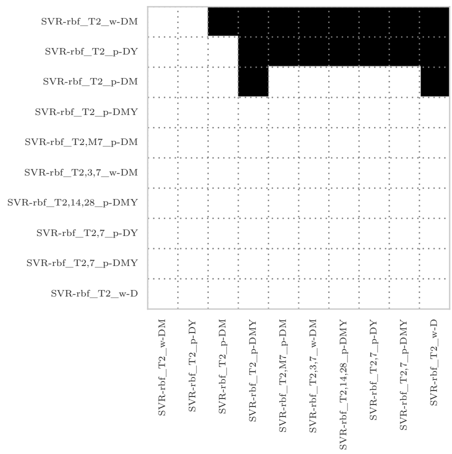


RMSE, DM test with $p = 2$

r SVR-poly-4

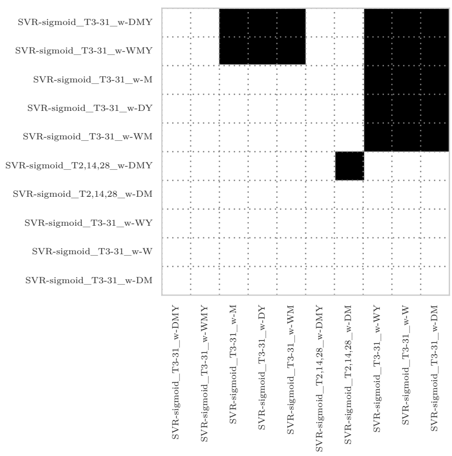


MAE, DM test with $p = 1$

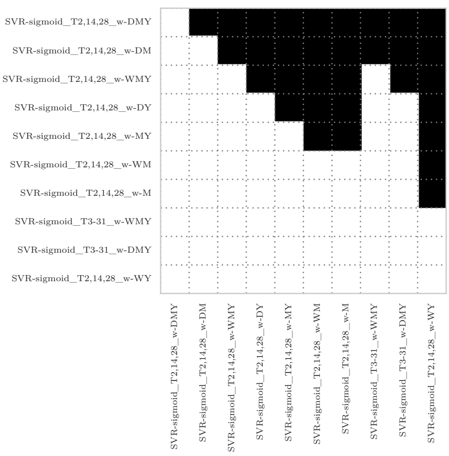


RMSE, DM test with $p = 2$

s SVR-rbf



MAE, DM test with $p = 1$



RMSE, DM test with $p = 2$

t SVR-sigmoid

Figure C.1: The results for DM tests for selections of individual estimators by individual measures. A filled square at $[i, j]$ denotes that the forecast of estimator i is better than the forecast j by the DM test on 5% significance level.

C.2 Summaries

rRMSE	Var.	W.	RMSE	MAE	rMAE	rPareto
1/2741	T2,7,14,28,M7,M30-W	No	**8.6204**	6.3587	3/2716	1/264
2/2744	T2,7,14,28,M7,M30-WM	No	**8.6234**	6.3443	1/2711	1/264
3/2747	T2,7,14,28,M7,M30-WY	No	**8.6258**	6.3497	2/2713	2/265
4/2756	T2,7,M7-WY	No	**8.6317**	6.5221	52/2852	3/274
5/2759	T2,M7,30-WMY	No	**8.6372**	6.3880	6/2748	3/268
6/2761	T2,M7-WY	No	**8.6417**	6.5421	64/2865	4/275
7/2764	T2,M7,30-W	No	**8.6447**	6.3980	10/2762	4/269
8/2767	T2,7,M7-WMY	No	**8.6477**	6.5339	59/2859	5/275
9/2772	T2,M7,30-WM	No	**8.6506**	6.3935	8/2756	4/269
10/2773	T2,M7-W	No	**8.6519**	6.5387	62/2862	6/276

Table C.1: The list of 10 best AB-DT estimators by RMSE. Column *rRMSE* contain the rank by RMSE within the class and also the overal rank (equivalety for *rMAE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rMAE	Var.	W.	RMSE	MAE	rRMSE	rPareto
1/2711	T2,7,14,28,M7,M30-WM	No	8.6234	**6.3443**	2/2744	1/264
2/2713	T2,7,14,28,M7,M30-WY	No	8.6258	**6.3497**	3/2747	2/265
3/2716	T2,7,14,28,M7,M30-W	No	8.6204	**6.3587**	1/2741	1/264
4/2735	T2,7,14,28,M7,M30-WMY	No	8.6572	**6.3752**	12/2775	3/267
5/2738	T2,7,14,28,M7,M30-WY	Yes	8.7161	**6.3789**	17/2795	4/270
6/2748	T2,M7,30-WMY	No	8.6372	**6.3880**	5/2759	3/268
7/2752	T2,7,14,28,M7,M30-W	Yes	8.7190	**6.3915**	18/2798	5/271
8/2756	T2,M7,30-WM	No	8.6506	**6.3935**	9/2772	4/269
9/2759	T2,7,14,28,M7,M30-WM	Yes	8.7334	**6.3959**	19/2805	6/272
10/2762	T2,M7,30-W	No	8.6447	**6.3980**	7/2764	4/269

Table C.2: The list of 10 best AB-DT estimators by MAE. Column *rMAE* contain the rank by MAE within the class and also the overal rank (equivalety for *rRMSE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rRMSE	**Var.**	**W.**	**RMSE**	**MAE**	**rMAE**	**rPareto**
1/157	T2,M7,30-D	No	**7.4500**	5.4696	1/252	1/38
2/222	T2,M7-D	No	**7.4969**	5.5392	2/468	2/45
3/224	T2,7,M7-D	No	**7.4972**	5.5454	3/497	3/47
4/238	T2,7,14,28,M7,M30-D	No	**7.5081**	5.5539	9/540	4/48
5/298	T2,7,M7-DY	No	**7.5449**	5.5841	15/669	5/56
6/299	T2,M7,30-DM	No	**7.5451**	5.5735	11/619	5/56
7/311	T2,M7-DY	No	**7.5498**	5.5841	14/668	6/58
8/329	T2,14,28-DM	No	**7.5579**	5.6268	21/868	7/60
9/368	T2,M7-D	Yes	**7.5754**	5.5950	16/708	7/61
10/382	T2-DM	Yes	**7.5808**	5.6154	20/805	8/62

Table C.3: The list of 10 best AB-LR estimators by RMSE. Column *rRMSE* contain the rank by RMSE within the class and also the overal rank (equivalety for *rMAE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rMAE	**Var.**	**W.**	**RMSE**	**MAE**	**rRMSE**	**rPareto**
1/252	T2,M7,30-D	No	7.4500	**5.4696**	1/157	1/38
2/468	T2,M7-D	No	7.4969	**5.5392**	2/222	2/45
3/497	T2,7,M7-D	No	7.4972	**5.5454**	3/224	3/47
4/503	T2-DMY	No	7.5834	**5.5466**	12/391	4/62
5/530	T2,3,7-DMY	No	7.5848	**5.5516**	13/396	5/64
6/531	T2,7-DMY	No	7.5859	**5.5518**	14/400	6/65
7/538	T2,7-DMY	Yes	7.5981	**5.5527**	17/436	7/68
8/539	T2-DMY	Yes	7.5963	**5.5539**	16/427	7/68
9/540	T2,7,14,28,M7,M30-D	No	7.5081	**5.5539**	4/238	4/48
10/578	T2,M7,30-D	Yes	7.6107	**5.5635**	21/490	8/70

Table C.4: The list of 10 best AB-LR estimators by MAE. Column *rMAE* contain the rank by MAE within the class and also the overal rank (equivalety for *rRMSE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rRMSE	**Var.**	**W.**	**RMSE**	**MAE**	**rMAE**	**rPareto**
1/392	T2,M7-D	No	**7.5839**	5.5672	8/599	1/63
2/404	T2,M7,30-D	Yes	**7.5874**	5.4947	1/308	1/52
3/411	T2,M7-DY	No	**7.5895**	5.5604	7/567	2/66
4/454	T2,M7,30-DY	No	**7.6020**	5.5170	2/367	2/59
5/479	T2-DMY	Yes	**7.6082**	5.5564	5/550	3/69
6/492	T2,7,M7-D	No	**7.6109**	5.5822	14/660	4/72
7/522	T2,M7-D	Yes	**7.6186**	5.5483	4/512	3/70
8/552	T2,7,14,28,M7,M30-D	No	**7.6289**	5.5583	6/556	4/74
9/574	T2,7-DMY	No	**7.6360**	5.5751	11/627	5/77
10/601	T2,M7,30-DM	No	**7.6437**	5.5884	15/684	6/80

Table C.5: The list of 10 best ANN48-0.5d-linearestimators by RMSE. Column *rRMSE* contain the rank by RMSE within the class and also the overal rank (equivalety for *rMAE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rMAE	Var.	W.	RMSE	MAE	rRMSE	rPareto
1/308	T2,M7,30-D	Yes	7.5874	**5.4947**	2/404	1/52
2/367	T2,M7,30-DY	No	7.6020	**5.5170**	4/454	2/59
3/431	T2,7,14,28,M7,M30-D	Yes	7.6542	**5.5313**	15/640	3/67
4/512	T2,M7-D	Yes	7.6186	**5.5483**	7/522	3/70
5/550	T2-DMY	Yes	7.6082	**5.5564**	5/479	3/69
6/556	T2,7,14,28,M7,M30-D	No	7.6289	**5.5583**	8/552	4/74
7/567	T2,M7-DY	No	7.5895	**5.5604**	3/411	2/66
8/599	T2,M7-D	No	7.5839	**5.5672**	1/392	1/63
9/602	T2,M7,30-D	No	7.6480	**5.5685**	11/615	5/77
10/614	T2,M7,30-DY	Yes	7.7331	**5.5719**	37/932	6/88

Table C.6: The list of 10 best ANN48-0.5d-linear estimators by MAE. Column *rMAE* contain the rank by MAE within the class and also the overal rank (equivalety for *rRMSE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rRMSE	Var.	W.	RMSE	MAE	rMAE	rPareto
1/677	T3-31-DM	No	**7.6662**	5.6894	25/1230	1/90
2/790	T2,M7,30-D	No	**7.6979**	5.6188	9/820	1/92
3/828	T2,7,14,28,M7,M30-DMY	No	**7.7107**	5.6548	17/1034	2/99
4/832	T2,M7,30-DMY	Yes	**7.7115**	5.5598	1/563	1/82
5/843	T2,7,14,28,M7,M30-DM	No	**7.7140**	5.6848	22/1201	3/100
6/868	T2,M7,30-D	Yes	**7.7199**	5.6070	6/765	2/92
7/884	T2,M7,30-DY	No	**7.7235**	5.6364	14/929	3/99
8/887	T2-7-DM	Yes	**7.7241**	5.5664	2/595	2/86
9/917	T2,7,14,28,M7,M30-DMY	Yes	**7.7301**	5.5996	5/730	3/89
10/923	T2,M7,30-DM	No	**7.7314**	5.6796	21/1173	4/107

Table C.7: The list of 10 best ANN84-0.5d-1200e estimators by RMSE. Column *rRMSE* contain the rank by RMSE within the class and also the overal rank (equivalety for *rMAE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rMAE	Var.	W.	RMSE	MAE	rRMSE	rPareto
1/563	T2,M7,30-DMY	Yes	7.7115	**5.5598**	4/832	1/82
2/595	T2-7-DM	Yes	7.7241	**5.5664**	8/887	2/86
3/635	T2-7-WMY	Yes	7.7676	**5.5767**	19/1076	3/91
4/717	T2,M7,30-DY	Yes	7.7474	**5.5963**	13/987	3/93
5/730	T2,7,14,28,M7,M30-DMY	Yes	7.7301	**5.5996**	9/917	3/89
6/765	T2,M7,30-D	Yes	7.7199	**5.6070**	6/868	2/92
7/814	T2-7-DY	Yes	7.8484	**5.6168**	54/1371	4/102
8/817	T2-7-D	Yes	7.8088	**5.6169**	33/1237	4/102
9/820	T2,M7,30-D	No	7.6979	**5.6188**	2/790	1/92
10/859	T2-7-WM	Yes	7.8020	**5.6253**	28/1210	4/107

Table C.8: The list of 10 best ANN84-0.5d-1200e estimators by MAE. Column *rMAE* contain the rank by MAE within the class and also the overal rank (equivalety for *rRMSE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rRMSE	Var.	W.	RMSE	MAE	rMAE	rPareto
1/54	T2,7-DMY	Yes	**7.3999**	5.3492	1/61	1/11
2/57	T2,7,M7-DM	No	**7.4006**	5.4045	5/131	2/15
3/77	T2,7-DM	Yes	**7.4109**	5.3806	2/90	2/16
4/79	T2,M7,30-D	No	**7.4110**	5.4356	15/187	3/18
5/100	T2,7,M7-D	No	**7.4250**	5.4445	16/197	4/24
6/102	T2,7,M7-D	Yes	**7.4266**	5.3933	4/119	3/24
7/125	T2,7,M7-DY	No	**7.4381**	5.4259	12/171	4/30
8/130	T2,7,M7-DMY	No	**7.4400**	5.4114	7/145	4/27
9/135	T2,M7,30-D	Yes	**7.4427**	5.4173	10/155	5/28
10/141	T2,M7,30-DM	No	**7.4469**	5.4518	17/205	6/32

Table C.9: The list of 10 best BR estimators by RMSE. Column *rRMSE* contain the rank by RMSE within the class and also the overal rank (equivalety for *rMAE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rMAE	Var.	W.	RMSE	MAE	rRMSE	rPareto
1/61	T2,7-DMY	Yes	7.3999	**5.3492**	1/54	1/11
2/90	T2,7-DM	Yes	7.4109	**5.3806**	3/77	2/16
3/108	T2,7,M7-DY	Yes	7.4488	**5.3874**	13/152	3/25
4/119	T2,7,M7-D	Yes	7.4266	**5.3933**	6/102	3/24
5/131	T2,7,M7-DM	No	7.4006	**5.4045**	2/57	2/15
6/143	T2,7-DY	Yes	7.4657	**5.4103**	17/180	4/29
7/145	T2,7,M7-DMY	No	7.4400	**5.4114**	8/130	4/27
8/146	T2,7,M7-DMY	Yes	7.4847	**5.4118**	20/202	5/30
9/148	T2,7,M7-DM	Yes	7.4569	**5.4127**	16/167	5/29
10/155	T2,M7,30-D	Yes	7.4427	**5.4173**	9/135	5/28

Table C.10: The list of 10 best BR estimators by MAE. Column *rMAE* contain the rank by MAE within the class and also the overal rank (equivalety for *rRMSE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rRMSE	Var.	W.	RMSE	MAE	rMAE	rPareto
1/7	T2,7,M7-DM	Yes	**7.3578**	5.3083	6/22	1/4
2/11	T2,7,M7-DMY	Yes	**7.3662**	5.2980	2/12	1/4
3/14	T2,7-DMY	Yes	**7.3728**	5.3382	13/46	2/6
4/23	T2,3,7-DMY	Yes	**7.3794**	5.3071	5/20	2/6
5/24	T2,7,M7-DM	No	**7.3796**	5.3568	19/69	3/8
6/29	T2,7,M7-DMY	No	**7.3805**	5.3473	17/58	3/8
7/34	T2,M7,30-DM	Yes	**7.3839**	5.3051	4/17	2/5
8/37	T2,7,14,28,M7,M30-DM	Yes	**7.3854**	5.3173	8/29	3/8
9/38	T2,7,14,28,M7,M30-DMY	Yes	**7.3855**	5.3021	3/15	2/6
10/39	T2,M7,30-DMY	Yes	**7.3877**	5.2898	1/6	1/3

Table C.11: The list of 10 best EN estimators by RMSE. Column *rRMSE* contain the rank by RMSE within the class and also the overal rank (equivalety for *rMAE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rMAE	Var.	W.	RMSE	MAE	rRMSE	rPareto
1/6	T2,M7,30-DMY	Yes	7.3877	**5.2898**	10/39	1/3
2/12	T2,7,M7-DMY	Yes	7.3662	**5.2980**	2/11	1/4
3/15	T2,7,14,28,M7,M30-DMY	Yes	7.3855	**5.3021**	9/38	2/6
4/17	T2,M7,30-DM	Yes	7.3839	**5.3051**	7/34	2/5
5/20	T2,3,7-DMY	Yes	7.3794	**5.3071**	4/23	2/6
6/22	T2,7,M7-DM	Yes	7.3578	**5.3083**	1/7	1/4
7/28	T2,M7,30-DY	Yes	7.4059	**5.3170**	16/66	3/9
8/29	T2,7,14,28,M7,M30-DM	Yes	7.3854	**5.3173**	8/37	3/8
9/34	T2,7,14,28,M7,M30-DY	Yes	7.4103	**5.3246**	20/75	4/11
10/35	T2-7-DMY	Yes	7.4399	**5.3265**	34/129	5/12

Table C.12: The list of 10 best EN estimators by MAE. Column *rMAE* contain the rank by MAE within the class and also the overal rank (equivalety for *rRMSE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rRMSE	Var.	W.	RMSE	MAE	rMAE	rPareto
1/275	T2,M7,30-DY	No	**7.5334**	5.4693	2/251	1/46
2/294	T2,M7,30-DM	No	**7.5435**	5.4862	5/286	2/47
3/296	T2,7,14,28,M7,M30-DM	No	**7.5441**	5.5494	21/518	3/55
4/313	T2,7,14,28,M7,M30-DY	No	**7.5512**	5.5340	17/442	3/55
5/342	T2,7,14,28,M7,M30-DMY	No	**7.5652**	5.5363	18/450	4/57
6/352	T2,M7,30-DMY	No	**7.5692**	5.4750	3/263	2/49
7/358	T2,7,14,28,M7,M30-D	Yes	**7.5718**	5.5160	9/362	3/56
8/363	T2,7,14,28,M7,M30-D	No	**7.5740**	5.5786	31/645	5/60
9/364	T2,M7,30-DY	Yes	**7.5741**	5.4644	1/243	1/44
10/366	T2,7,14,28,M7,M30-DM	Yes	**7.5751**	5.5240	12/399	4/58

Table C.13: The list of 10 best KRR-linear estimators by RMSE. Column *rRMSE* contain the rank by RMSE within the class and also the overal rank (equivalety for *rMAE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rMAE	Var.	W.	RMSE	MAE	rRMSE	rPareto
1/243	T2,M7,30-DY	Yes	7.5741	**5.4644**	9/364	1/44
2/251	T2,M7,30-DY	No	7.5334	**5.4693**	1/275	1/46
3/263	T2,M7,30-DMY	No	7.5692	**5.4750**	6/352	2/49
4/274	T2,M7,30-DMY	Yes	7.6152	**5.4821**	19/507	3/53
5/286	T2,M7,30-DM	No	7.5435	**5.4862**	2/294	2/47
6/316	T2,7,14,28,M7,M30-DY	Yes	7.5758	**5.4971**	11/370	3/52
7/323	T2-7-DMY	Yes	7.6300	**5.4986**	26/556	4/57
8/329	T2,M7,30-DM	Yes	7.6073	**5.5003**	18/476	4/56
9/362	T2,7,14,28,M7,M30-D	Yes	7.5718	**5.5160**	7/358	3/56
10/390	T2,7,14,28,M7,M30-DMY	Yes	7.5983	**5.5215**	17/440	4/61

Table C.14: The list of 10 best KRR-linear estimators by MAE. Column *rMAE* contain the rank by MAE within the class and also the overal rank (equivalety for *rRMSE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rRMSE	Var.	W.	RMSE	MAE	rMAE	rPareto
1/123	T2,7-DM	No	**7.4358**	5.4503	1/204	1/30
2/165	T2,7-D	No	**7.4564**	5.4760	9/264	2/39
3/185	T2,14,28-DM	No	**7.4696**	5.5265	23/415	3/41
4/191	T2,14,28-DMY	No	**7.4720**	5.4965	12/315	3/42
5/193	T2,7-DMY	No	**7.4741**	5.4550	2/216	2/38
6/214	T2,7,M7-DM	No	**7.4935**	5.4613	3/236	3/41
7/218	T2,7-DY	No	**7.4951**	5.4673	7/249	4/42
8/230	T2,7,M7-D	No	**7.5043**	5.4761	10/265	5/43
9/267	T2,7,M7-DMY	No	**7.5309**	5.4658	5/247	4/44
10/271	T2,7-WM	No	**7.5323**	5.4672	6/248	5/45

Table C.15: The list of 10 best KRR-poly-2 estimators by RMSE. Column *rRMSE* contain the rank by RMSE within the class and also the overal rank (equivalety for *rMAE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rMAE	Var.	W.	RMSE	MAE	rRMSE	rPareto
1/204	T2,7-DM	No	7.4358	**5.4503**	1/123	1/30
2/216	T2,7-DMY	No	7.4741	**5.4550**	5/193	2/38
3/236	T2,7,M7-DM	No	7.4935	**5.4613**	6/214	3/41
4/245	T2,7-WMY	No	7.5388	**5.4645**	11/282	4/44
5/247	T2,7,M7-DMY	No	7.5309	**5.4658**	9/267	4/44
6/248	T2,7-WM	No	7.5323	**5.4672**	10/271	5/45
7/249	T2,7-DY	No	7.4951	**5.4673**	7/218	4/42
8/259	T2,7,M7-DY	No	7.5443	**5.4739**	12/297	6/47
9/264	T2,7-D	No	7.4564	**5.4760**	2/165	2/39
10/265	T2,7,M7-D	No	7.5043	**5.4761**	8/230	5/43

Table C.16: The list of 10 best KRR-poly-2 estimators by MAE. Column *rMAE* contain the rank by MAE within the class and also the overal rank (equivalety for *rRMSE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rRMSE	Var.	W.	RMSE	MAE	rMAE	rPareto
1/3142	nl-DMY	No	**9.1701**	6.7607	1/3070	1/321
1/3142	nl-DMY	Yes	**9.1701**	6.7607	1/3070	1/321
3/3232	nl-WMY	No	**9.3553**	6.9675	3/3232	2/337
3/3232	nl-WMY	Yes	**9.3553**	6.9675	3/3232	2/337
5/3289	nl-MY	No	**9.4699**	7.0923	5/3360	3/346
5/3289	nl-MY	Yes	**9.4699**	7.0923	5/3360	3/346
7/3439	T2-MY	No	**9.7349**	7.3894	7/3574	4/368
7/3439	T2,7-MY	No	**9.7349**	7.3894	7/3574	4/368
9/3533	T2-DMY	No	**9.8209**	7.4871	9/3692	5/383
9/3533	T2-DMY	Yes	**9.8209**	7.4871	9/3692	5/383
9/3533	T2,7-DMY	No	**9.8209**	7.4871	9/3692	5/383
9/3533	T2,7-DMY	Yes	**9.8209**	7.4871	9/3692	5/383

Table C.17: The list of 10 best Lars estimators by RMSE. Column *rRMSE* contain the rank by RMSE within the class and also the overal rank (equivalety for *rMAE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rMAE	**Var.**	**W.**	**RMSE**	**MAE**	**rRMSE**	**rPareto**
1/3070	nl-DMY	No	9.1701	**6.7607**	1/3142	1/321
1/3070	nl-DMY	Yes	9.1701	**6.7607**	1/3142	1/321
3/3232	nl-WMY	No	9.3553	**6.9675**	3/3232	2/337
3/3232	nl-WMY	Yes	9.3553	**6.9675**	3/3232	2/337
5/3360	nl-MY	No	9.4699	**7.0923**	5/3289	3/346
5/3360	nl-MY	Yes	9.4699	**7.0923**	5/3289	3/346
7/3574	T2-MY	No	9.7349	**7.3894**	7/3439	4/368
7/3574	T2,7-MY	No	9.7349	**7.3894**	7/3439	4/368
9/3692	T2-DMY	No	9.8209	**7.4871**	9/3533	5/383
9/3692	T2-DMY	Yes	9.8209	**7.4871**	9/3533	5/383
9/3692	T2,7-DMY	No	9.8209	**7.4871**	9/3533	5/383
9/3692	T2,7-DMY	Yes	9.8209	**7.4871**	9/3533	5/383

Table C.18: The list of 10 best Lars estimators by MAE. Column *rMAE* contain the rank by MAE within the class and also the overal rank (equivalety for *rRMSE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rRMSE	**Var.**	**W.**	**RMSE**	**MAE**	**rMAE**	**rPareto**
1/6	T2,7,M7-DM	Yes	**7.3566**	5.3081	6/21	1/3
2/10	T2,7,M7-DMY	Yes	**7.3647**	5.2977	2/10	1/3
3/17	T2,7,M7-DM	No	**7.3758**	5.3554	21/68	2/7
4/21	T2,3,7-DMY	Yes	**7.3785**	5.3057	5/19	2/5
5/22	T2,7-DMY	Yes	**7.3791**	5.3510	19/62	3/7
6/30	T2,7,M7-DMY	No	**7.3816**	5.3490	18/60	3/9
7/33	T2,M7,30-DM	Yes	**7.3838**	5.3052	4/18	2/5
8/35	T2,7,14,28,M7,M30-DMY	Yes	**7.3848**	5.3014	3/14	2/5
9/36	T2,7,14,28,M7,M30-DM	Yes	**7.3848**	5.3166	8/26	3/7
10/42	T2,M7,30-DMY	Yes	**7.3896**	5.2937	1/8	1/4

Table C.19: The list of 10 best Lasso estimators by RMSE. Column *rRMSE* contain the rank by RMSE within the class and also the overal rank (equivalety for *rMAE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rMAE	**Var.**	**W.**	**RMSE**	**MAE**	**rRMSE**	**rPareto**
1/8	T2,M7,30-DMY	Yes	7.3896	**5.2937**	10/42	1/4
2/10	T2,7,M7-DMY	Yes	7.3647	**5.2977**	2/10	1/3
3/14	T2,7,14,28,M7,M30-DMY	Yes	7.3848	**5.3014**	8/35	2/5
4/18	T2,M7,30-DM	Yes	7.3838	**5.3052**	7/33	2/5
5/19	T2,3,7-DMY	Yes	7.3785	**5.3057**	4/21	2/5
6/21	T2,7,M7-DM	Yes	7.3566	**5.3081**	1/6	1/3
7/23	T2-7-DMY	Yes	7.4069	**5.3113**	19/69	3/8
8/26	T2,7,14,28,M7,M30-DM	Yes	7.3848	**5.3166**	9/36	3/7
9/27	T2,M7,30-DY	Yes	7.4050	**5.3166**	17/65	4/8
10/33	T2,7,14,28,M7,M30-DY	Yes	7.4091	**5.3234**	21/74	5/10

Table C.20: The list of 10 best Lasso estimators by MAE. Column *rMAE* contain the rank by MAE within the class and also the overal rank (equivalety for *rRMSE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rRMSE	**Var.**	**W.**	**RMSE**	**MAE**	**rMAE**	**rPareto**
1/1	T2,7,M7-DM	Yes	**7.3363**	5.2980	8/11	1/1
2/2	T2,M7,30-DM	Yes	**7.3465**	5.2829	4/4	1/1
3/3	T2,7,14,28,M7,M30-DM	Yes	**7.3491**	5.2814	3/3	1/1
4/4	T2,7,M7-DMY	Yes	**7.3533**	5.2887	5/5	2/2
5/5	T2,7,14,28,M7,M30-DMY	Yes	**7.3558**	5.2710	1/1	1/1
6/8	T2,7,M7-DM	No	**7.3596**	5.3405	21/49	3/5
7/9	T2,7-DM	Yes	**7.3613**	5.3367	18/43	3/5
8/12	T2,3,7-DMY	Yes	**7.3687**	5.2946	7/9	3/3
9/13	T2,M7,30-DMY	Yes	**7.3689**	5.2763	2/2	2/2
10/15	T2,7,M7-DMY	No	**7.3746**	5.3356	17/42	4/5

Table C.21: The list of 10 best LassoLars estimators by RMSE. Column *rRMSE* contain the rank by RMSE within the class and also the overal rank (equivalety for *rMAE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rMAE	**Var.**	**W.**	**RMSE**	**MAE**	**rRMSE**	**rPareto**
1/1	T2,7,14,28,M7,M30-DMY	Yes	7.3558	**5.2710**	5/5	1/1
2/2	T2,M7,30-DMY	Yes	7.3689	**5.2763**	9/13	2/2
3/3	T2,7,14,28,M7,M30-DM	Yes	7.3491	**5.2814**	3/3	1/1
4/4	T2,M7,30-DM	Yes	7.3465	**5.2829**	2/2	1/1
5/5	T2,7,M7-DMY	Yes	7.3533	**5.2887**	4/4	2/2
6/7	T2-7-DMY	Yes	7.3798	**5.2931**	16/26	3/3
7/9	T2,3,7-DMY	Yes	7.3687	**5.2946**	8/12	3/3
8/11	T2,7,M7-DM	Yes	7.3363	**5.2980**	1/1	1/1
9/13	T2,7,14,28,M7,M30-DY	Yes	7.3896	**5.3012**	23/43	4/5
10/16	T2,M7,30-DY	Yes	7.3923	**5.3023**	25/45	5/7

Table C.22: The list of 10 best LassoLars estimators by MAE. Column *rMAE* contain the rank by MAE within the class and also the overal rank (equivalety for *rRMSE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rRMSE	**Var.**	**W.**	**RMSE**	**MAE**	**rMAE**	**rPareto**
1/133	T2,7-DM	Yes	**7.4408**	5.4521	2/206	1/31
2/151	T2,7,M7-D	No	**7.4487**	5.4886	5/291	2/37
3/155	T2,7-DM	No	**7.4497**	5.4905	6/299	3/38
4/166	T2,M7,30-D	No	**7.4567**	5.4729	3/258	2/39
5/182	T2,7-DMY	Yes	**7.4673**	5.4445	1/198	1/35
6/184	T2,7-DMY	No	**7.4689**	5.4784	4/271	3/40
7/210	T2,3,7-DM	No	**7.4903**	5.5182	9/371	4/44
8/223	T2,M7-D	No	**7.4972**	5.5406	14/475	5/46
9/226	T2,14,28-DM	No	**7.4997**	5.5784	27/641	6/48
10/228	T2,3,7-DMY	No	**7.5012**	5.5108	7/344	4/45

Table C.23: The list of 10 best OLS estimators by RMSE. Column *rRMSE* contain the rank by RMSE within the class and also the overal rank (equivalety for *rMAE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rMAE	Var.	W.	RMSE	MAE	rRMSE	rPareto
1/198	T2,7-DMY	Yes	7.4673	**5.4445**	5/182	1/35
2/206	T2,7-DM	Yes	7.4408	**5.4521**	1/133	1/31
3/258	T2,M7,30-D	No	7.4567	**5.4729**	4/166	2/39
4/271	T2,7-DMY	No	7.4689	**5.4784**	6/184	3/40
5/291	T2,7,M7-D	No	7.4487	**5.4886**	2/151	2/37
6/299	T2,7-DM	No	7.4497	**5.4905**	3/155	3/38
7/344	T2,3,7-DMY	No	7.5012	**5.5108**	10/228	4/45
8/368	T2,7-DY	Yes	7.5373	**5.5175**	17/281	5/51
9/371	T2,3,7-DM	No	7.4903	**5.5182**	7/210	4/44
10/402	T2,7,M7-D	Yes	7.5305	**5.5244**	12/266	5/50

Table C.24: The list of 10 best OLS estimators by MAE. Column *rMAE* contain the rank by MAE within the class and also the overal rank (equivalety for *rRMSE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rRMSE	Var.	W.	RMSE	MAE	rMAE	rPareto
1/2425	T2-DY	No	**8.2471**	6.1772	36/2549	1/228
1/2425	T2,7-DY	No	**8.2471**	6.1772	36/2549	1/228
3/2442	T2,14,28-D	No	**8.2574**	6.0914	1/2391	1/225
4/2450	T2,14,28-DY	No	**8.2641**	6.0984	3/2409	2/226
5/2454	T2,14,28-W	No	**8.2651**	6.1028	5/2427	3/228
6/2460	T2,14,28-DM	No	**8.2718**	6.1146	13/2449	4/229
7/2462	T2,14,28-WM	No	**8.2727**	6.1084	10/2442	4/229
8/2470	T2,14,28-WMY	No	**8.2794**	6.1103	12/2444	5/230
9/2472	T2,14,28-WY	No	**8.2816**	6.1148	14/2450	6/231
10/2479	T2,14,28-DMY	No	**8.2854**	6.1039	7/2429	4/229

Table C.25: The list of 10 best RF estimators by RMSE. Column *rRMSE* contain the rank by RMSE within the class and also the overal rank (equivalety for *rMAE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rMAE	Var.	W.	RMSE	MAE	rRMSE	rPareto
1/2391	T2,14,28-D	No	8.2574	**6.0914**	3/2442	1/225
2/2400	T2-7-W	No	8.4305	**6.0953**	48/2617	2/229
3/2409	T2,14,28-DY	No	8.2641	**6.0984**	4/2450	2/226
4/2420	T2-7-WMY	No	8.4214	**6.1010**	46/2612	3/231
5/2427	T2,14,28-W	No	8.2651	**6.1028**	5/2454	3/228
6/2428	T2-7-WY	No	8.4355	**6.1033**	53/2626	4/232
7/2429	T2,14,28-DMY	No	8.2854	**6.1039**	10/2479	4/229
8/2433	T2-7-D	No	8.4334	**6.1055**	52/2624	5/232
9/2440	T2-7-DM	No	8.4477	**6.1072**	59/2642	6/233
10/2442	T2,14,28-WM	No	8.2727	**6.1084**	7/2462	4/229

Table C.26: The list of 10 best RF estimators by MAE. Column *rMAE* contain the rank by MAE within the class and also the overal rank (equivalety for *rRMSE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rRMSE	Var.	W.	RMSE	MAE	rMAE	rPareto
1/136	T2,M7,30-D	No	**7.4435**	5.3998	2/127	1/25
2/208	T2,M7,30-DM	No	**7.4890**	5.4308	5/177	2/34
3/227	T2,7,M7-DM	No	**7.5011**	5.4831	9/277	3/44
4/233	T2,M7-DM	No	**7.5065**	5.4987	13/324	4/46
5/242	T2,M7,30-DY	No	**7.5117**	5.3953	1/120	1/28
6/255	T2,M7-DY	No	**7.5232**	5.5055	16/336	5/48
7/257	T2,M7,30-D	Yes	**7.5238**	5.4158	4/153	2/32
8/263	T2,M7-D	No	**7.5278**	5.5309	30/430	6/51
9/270	T2,7,M7-D	No	**7.5318**	5.5354	31/445	7/52
10/276	T2,M7,30-DMY	No	**7.5339**	5.4691	8/250	3/46

Table C.27: The list of 10 best SVR-linear-e estimators by RMSE. Column *rRMSE* contain the rank by RMSE within the class and also the overal rank (equivalety for *rMAE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rMAE	Var.	W.	RMSE	MAE	rRMSE	rPareto
1/120	T2,M7,30-DY	No	7.5117	**5.3953**	5/242	1/28
2/127	T2,M7,30-D	No	7.4435	**5.3998**	1/136	1/25
3/132	T2,M7,30-DY	Yes	7.5472	**5.4045**	12/308	2/30
4/153	T2,M7,30-D	Yes	7.5238	**5.4158**	7/257	2/32
5/177	T2,M7,30-DM	No	7.4890	**5.4308**	2/208	2/34
6/224	T2,M7,30-DMY	Yes	7.5919	**5.4567**	24/417	3/44
7/246	T2-7-DMY	Yes	7.6332	**5.4651**	40/565	4/48
8/250	T2,M7,30-DMY	No	7.5339	**5.4691**	10/276	3/46
9/277	T2,7,M7-DM	No	7.5011	**5.4831**	3/227	3/44
10/283	T2,M7,30-DM	Yes	7.5772	**5.4861**	22/373	4/50

Table C.28: The list of 10 best SVR-linear-e estimators by MAE. Column *rMAE* contain the rank by MAE within the class and also the overal rank (equivalety for *rRMSE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rRMSE	Var.	W.	RMSE	MAE	rMAE	rPareto
1/186	T2,M7,30-D	No	**7.4698**	5.4589	7/232	1/38
2/203	T2,M7,30-DY	No	**7.4848**	5.4391	2/194	1/34
3/219	T2,M7,30-DM	No	**7.4953**	5.4485	4/202	2/36
4/231	T2,7,M7-DM	No	**7.5054**	5.4890	10/295	3/45
5/248	T2,M7-DM	No	**7.5151**	5.4946	12/307	4/47
6/251	T2,7,M7-DY	No	**7.5187**	5.5110	16/345	5/48
7/253	T2,7,14,28,M7,M30-DM	No	**7.5206**	5.5354	28/446	6/50
8/256	T2,M7,30-D	Yes	**7.5237**	5.4476	3/200	2/39
9/261	T2,7,14,28,M7,M30-D	No	**7.5266**	5.5450	36/495	7/52
10/272	T2,M7-D	No	**7.5325**	5.5370	29/453	7/53

Table C.29: The list of 10 best SVR-linear-e2 estimators by RMSE. Column *rRMSE* contain the rank by RMSE within the class and also the overal rank (equivalety for *rMAE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rMAE	Var.	W.	RMSE	MAE	rRMSE	rPareto
1/193	T2,M7,30-DY	Yes	7.5412	**5.4376**	14/288	1/39
2/194	T2,M7,30-DY	No	7.4848	**5.4391**	2/203	1/34
3/200	T2,M7,30-D	Yes	7.5237	**5.4476**	8/256	2/39
4/202	T2,M7,30-DM	No	7.4953	**5.4485**	3/219	2/36
5/211	T2,M7,30-DMY	No	7.5369	**5.4539**	12/279	3/40
6/228	T2,M7,30-DMY	Yes	7.5814	**5.4583**	29/384	4/45
7/232	T2,M7,30-D	No	7.4698	**5.4589**	1/186	1/38
8/237	T2,M7,30-DM	Yes	7.5549	**5.4614**	19/322	4/42
9/285	T2-7-DMY	Yes	7.6146	**5.4862**	37/506	5/53
10/295	T2,7,M7-DM	No	7.5054	**5.4890**	4/231	3/45

Table C.30: The list of 10 best SVR-linear-e2 estimators by MAE. Column *rMAE* contain the rank by MAE within the class and also the overal rank (equivalety for *rRMSE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rRMSE	Var.	W.	RMSE	MAE	rMAE	rPareto
1/1953	T2-DM	No	**8.0146**	5.9466	1/2126	1/180
2/2342	T2,14,28-DM	No	**8.1882**	6.0786	5/2373	2/213
3/2346	T2,7-WY	No	**8.1896**	5.9930	3/2241	2/202
4/2378	T2,7-DY	No	**8.2076**	5.9673	2/2173	2/197
5/2402	T2,14,28-DMY	No	**8.2227**	6.0973	8/2404	3/223
6/2441	T2,14,28-WM	No	**8.2563**	6.1542	12/2500	4/229
7/2452	T2-DMY	No	**8.2641**	6.0855	7/2382	3/224
8/2487	T2,7-DM	No	**8.2891**	6.0831	6/2377	3/224
9/2497	T2-DY	No	**8.2965**	6.0383	4/2316	3/211
10/2500	T2,7-WM	Yes	**8.3014**	6.1642	15/2522	5/236

Table C.31: The list of 10 best SVR-poly-2estimators by RMSE. Column *rRMSE* contain the rank by RMSE within the class and also the overal rank (equivalety for *rMAE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rMAE	Var.	W.	RMSE	MAE	rRMSE	rPareto
1/2126	T2-DM	No	8.0146	**5.9466**	1/1953	1/180
2/2173	T2,7-DY	No	8.2076	**5.9673**	4/2378	2/197
3/2241	T2,7-WY	No	8.1896	**5.9930**	3/2346	2/202
4/2316	T2-DY	No	8.2965	**6.0383**	9/2497	3/211
5/2373	T2,14,28-DM	No	8.1882	**6.0786**	2/2342	2/213
6/2377	T2,7-DM	No	8.2891	**6.0831**	8/2487	3/224
7/2382	T2-DMY	No	8.2641	**6.0855**	7/2452	3/224
8/2404	T2,14,28-DMY	No	8.2227	**6.0973**	5/2402	3/223
9/2411	T2,7-DMY	No	8.3250	**6.0985**	15/2527	4/230
10/2480	T2,7-WMY	No	8.3816	**6.1358**	18/2585	5/236

Table C.32: The list of 10 best SVR-poly-2 estimators by MAE. Column *rMAE* contain the rank by MAE within the class and also the overal rank (equivalety for *rRMSE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rRMSE	Var.	W.	RMSE	MAE	rMAE	rPareto
1/2892	T2,7,M7-DY	Yes	**8.8580**	6.5185	1/2851	1/292
2/2917	T2,7,M7-WY	Yes	**8.8978**	6.5763	2/2883	2/295
3/2940	T2,7,M7-D	Yes	**8.9203**	6.5799	3/2885	3/299
4/2979	T2,7,M7-W	Yes	**8.9586**	6.6344	6/2952	4/304
5/2984	T2,7,M7-DMY	Yes	**8.9654**	6.6024	4/2911	4/302
6/2995	T2,7,M7-DM	Yes	**8.9811**	6.6288	5/2945	5/303
7/3015	T2,7,M7-WMY	Yes	**9.0172**	6.6690	9/2988	6/308
8/3022	T2,7,M7-WM	Yes	**9.0291**	6.6890	13/3015	7/309
9/3052	T2,7-WM	Yes	**9.0722**	6.6758	11/2993	7/309
10/3060	T2,7-WY	Yes	**9.0790**	6.7025	16/3029	8/311

Table C.33: The list of 10 best SVR-poly-3 estimators by RMSE. Column *rRMSE* contain the rank by RMSE within the class and also the overal rank (equivalety for *rMAE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rMAE	Var.	W.	RMSE	MAE	rRMSE	rPareto
1/2851	T2,7,M7-DY	Yes	8.8580	**6.5185**	1/2892	1/292
2/2883	T2,7,M7-WY	Yes	8.8978	**6.5763**	2/2917	2/295
3/2885	T2,7,M7-D	Yes	8.9203	**6.5799**	3/2940	3/299
4/2911	T2,7,M7-DMY	Yes	8.9654	**6.6024**	5/2984	4/302
5/2945	T2,7,M7-DM	Yes	8.9811	**6.6288**	6/2995	5/303
6/2952	T2,7,M7-W	Yes	8.9586	**6.6344**	4/2979	4/304
7/2957	T2,7-DM	Yes	9.0946	**6.6405**	13/3074	6/312
8/2963	T2,7,M7-DY	No	9.1382	**6.6462**	22/3115	7/313
9/2988	T2,7,M7-WMY	Yes	9.0172	**6.6690**	7/3015	6/308
10/2990	T2,7-DY	Yes	9.1490	**6.6744**	23/3120	8/318

Table C.34: The list of 10 best SVR-poly-3 estimators by MAE. Column *rMAE* contain the rank by MAE within the class and also the overal rank (equivalety for *rRMSE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rRMSE	Var.	W.	RMSE	MAE	rMAE	rPareto
1/4289	T2,7,M7-DY	No	**10.6661**	7.6831	1/3979	1/437
2/4292	T2,7-DY	No	**10.6754**	7.8203	3/4114	2/456
3/4318	T2,7-DMY	No	**10.7516**	7.9046	7/4164	3/460
4/4382	T2,7,M7-WY	No	**10.8223**	7.8733	4/4145	3/459
5/4395	T2,7,M7-DMY	No	**10.8727**	7.8074	2/4105	2/452
6/4399	T2,7,14,28,M7,M30-WMY	Yes	**10.8916**	8.4583	45/4517	4/481
7/4408	T2,7,14,28,M7,M30-DMY	Yes	**10.9120**	8.4669	47/4532	5/482
8/4414	T2,7-Y	No	**10.9365**	8.1641	24/4369	4/477
9/4418	T2,7-DY	Yes	**10.9533**	7.9957	13/4259	4/466
10/4439	T2,7,14,28,M7,M30-WMY	No	**10.9900**	8.4491	44/4512	5/485

Table C.35: The list of 10 best SVR-poly-4 estimators by RMSE. Column *rRMSE* contain the rank by RMSE within the class and also the overal rank (equivalety for *rMAE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rMAE	Var.	W.	RMSE	MAE	rRMSE	rPareto
1/3979	T2,7,M7-DY	No	10.6661	**7.6831**	1/4289	1/437
2/4105	T2,7,M7-DMY	No	10.8727	**7.8074**	5/4395	2/452
3/4114	T2,7-DY	No	10.6754	**7.8203**	2/4292	2/456
4/4145	T2,7,M7-WY	No	10.8223	**7.8733**	4/4382	3/459
5/4157	T2,7,M7-DMY	Yes	11.1250	**7.8959**	21/4546	4/460
6/4158	T2,7,M7-D	No	11.1185	**7.8997**	19/4543	4/460
7/4164	T2,7-DMY	No	10.7516	**7.9046**	3/4318	3/460
8/4181	T2,7-WY	No	11.0025	**7.9200**	12/4447	4/463
9/4233	T2,7,M7-DY	Yes	11.1203	**7.9338**	20/4545	5/464
10/4240	T2,7-DMY	Yes	11.0202	**7.9541**	13/4451	5/466

Table C.36: The list of 10 best SVR-poly-4 estimators by MAE. Column *rMAE* contain the rank by MAE within the class and also the overal rank (equivalety for *rRMSE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rRMSE	Var.	W.	RMSE	MAE	rMAE	rPareto
1/337	T2-DM	Yes	**7.5631**	5.5590	12/559	1/57
2/487	T2-DY	No	**7.6096**	5.4941	3/306	1/55
3/578	T2-DM	No	**7.6371**	5.5419	11/481	2/72
4/703	T2-DMY	No	**7.6742**	5.5242	8/401	2/65
5/720	T2,M7-DM	No	**7.6809**	5.5355	9/447	3/69
6/735	T2,3,7-DM	Yes	**7.6855**	5.6467	50/976	4/91
7/742	T2,14,28-DMY	No	**7.6869**	5.5943	21/706	4/84
8/770	T2,7-DY	No	**7.6935**	5.5126	5/351	2/60
9/781	T2,7-DMY	No	**7.6959**	5.4543	1/212	1/41
10/813	T2-D	Yes	**7.7061**	5.6332	42/908	5/97

Table C.37: The list of 10 best SVR-rbf estimators by RMSE. Column *rRMSE* contain the rank by RMSE within the class and also the overal rank (equivalety for *rMAE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rMAE	Var.	W.	RMSE	MAE	rRMSE	rPareto
1/212	T2,7-DMY	No	7.6959	**5.4543**	9/781	1/41
2/270	T2,7-DM	No	7.7169	**5.4783**	12/855	2/52
3/306	T2-DY	No	7.6096	**5.4941**	2/487	1/55
4/350	T2,7-DMY	Yes	7.7611	**5.5125**	19/1049	3/60
5/351	T2,7-DY	No	7.6935	**5.5126**	8/770	2/60
6/361	T2,7-D	No	7.7732	**5.5159**	23/1101	4/62
7/374	T2,7-WMY	No	7.7579	**5.5189**	18/1041	3/62
8/401	T2-DMY	No	7.6742	**5.5242**	4/703	2/65
9/447	T2,M7-DM	No	7.6809	**5.5355**	5/720	3/69
10/469	T2,7-WM	No	7.7792	**5.5393**	25/1123	5/71

Table C.38: The list of 10 best SVR-rbf estimators by MAE. Column *rMAE* contain the rank by MAE within the class and also the overal rank (equivalety for *rRMSE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rRMSE	**Var.**	**W.**	**RMSE**	**MAE**	**rMAE**	**rPareto**
1/2959	T2,14,28-DMY	Yes	**8.9394**	6.6319	3/2951	1/303
2/2983	T2,14,28-DM	Yes	**8.9646**	6.6427	7/2960	2/305
3/2994	T2,14,28-WMY	Yes	**8.9786**	6.6567	13/2973	3/306
4/3007	T2,14,28-DY	Yes	**9.0105**	6.6768	19/2995	4/307
5/3021	T2,14,28-MY	Yes	**9.0288**	6.6906	25/3017	5/309
6/3032	T2,14,28-WM	Yes	**9.0427**	6.7110	31/3039	6/311
7/3038	T2,14,28-M	Yes	**9.0532**	6.7050	29/3031	6/311
8/3042	T3-31-WMY	Yes	**9.0637**	6.6173	2/2932	1/307
9/3043	T3-31-DMY	Yes	**9.0637**	6.6168	1/2930	1/307
10/3053	T2,14,28-WY	Yes	**9.0752**	6.7291	36/3050	7/313

Table C.39: The list of 10 best SVR-sigmoid estimators by RMSE. Column *rRMSE* contain the rank by RMSE within the class and also the overal rank (equivalety for *rMAE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

rMAE	**Var.**	**W.**	**RMSE**	**MAE**	**rRMSE**	**rPareto**
1/2930	T3-31-DMY	Yes	9.0637	**6.6168**	9/3043	1/307
2/2932	T3-31-WMY	Yes	9.0637	**6.6173**	8/3042	1/307
3/2951	T2,14,28-DMY	Yes	8.9394	**6.6319**	1/2959	1/303
4/2953	T3-31-M	Yes	9.0840	**6.6379**	11/3067	2/310
5/2955	T3-31-DY	Yes	9.0896	**6.6399**	13/3069	3/311
6/2958	T3-31-WM	Yes	9.0903	**6.6409**	14/3071	4/312
7/2960	T2,14,28-DM	Yes	8.9646	**6.6427**	2/2983	2/305
8/2964	T3-31-WY	Yes	9.0974	**6.6472**	16/3080	5/313
9/2965	T3-31-W	Yes	9.1000	**6.6483**	17/3086	6/314
10/2966	T3-31-DM	Yes	9.0921	**6.6485**	15/3073	5/313

Table C.40: The list of 10 best SVR-sigmoid estimators by MAE. Column *rMAE* contain the rank by MAE within the class and also the overal rank (equivalety for *rRMSE* and *rPareto*). The column *Var.* contains variables and dummies used by the estimator and the columnd *W.* shows whether the weather data were included.

Appendix D

Impact of weather data

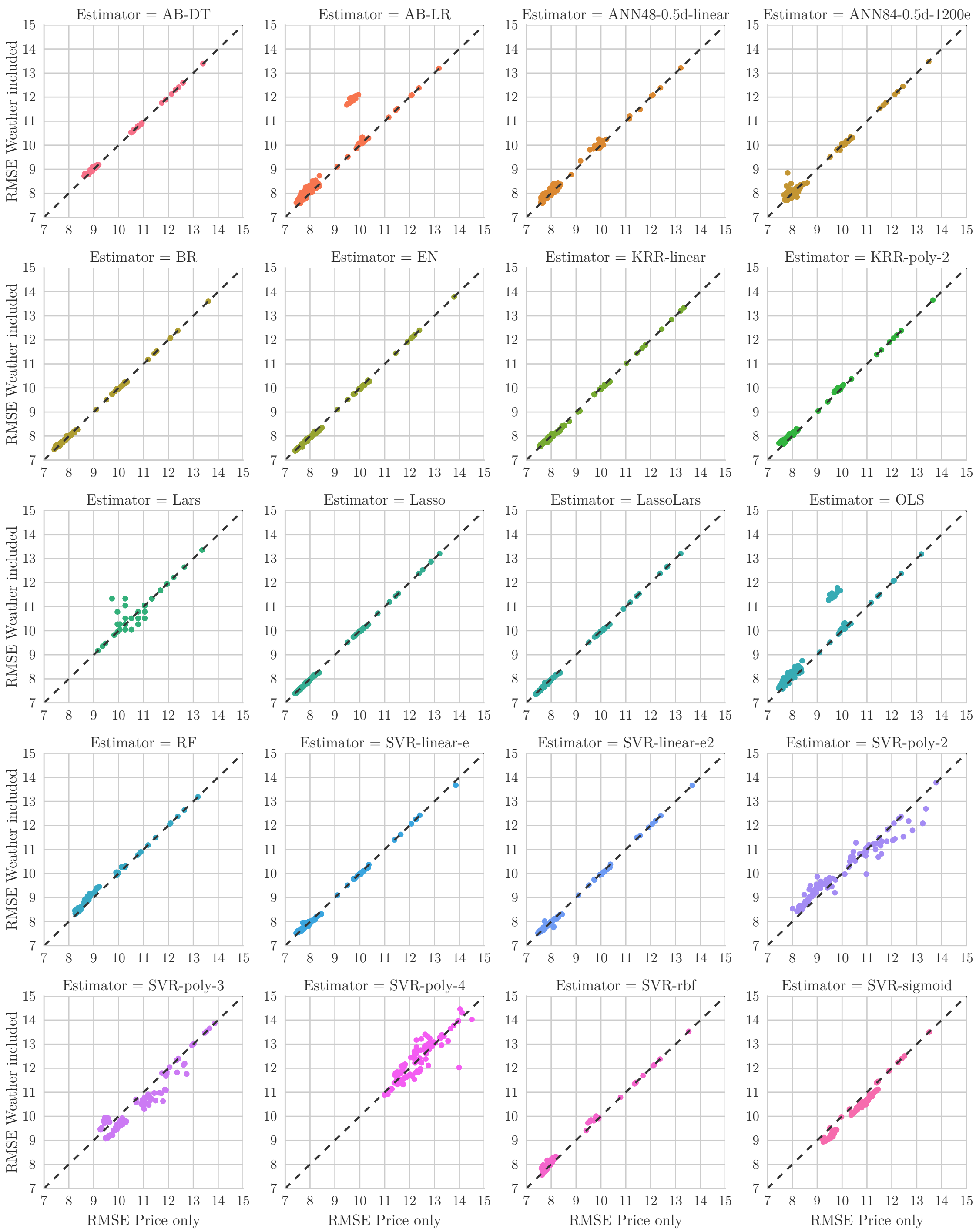


Figure D.1: The impact of weather data on RMSE. Each point represent a pair of estimators A and B that differ only by the inclusion of weather data. The x axis shows the RMSE of the estimator without the weather data and the y axis hows the RMSE of the estimator with the weather data. The closer the estimators are to the dashed line, the smaller influence have the weather data on the estimator.

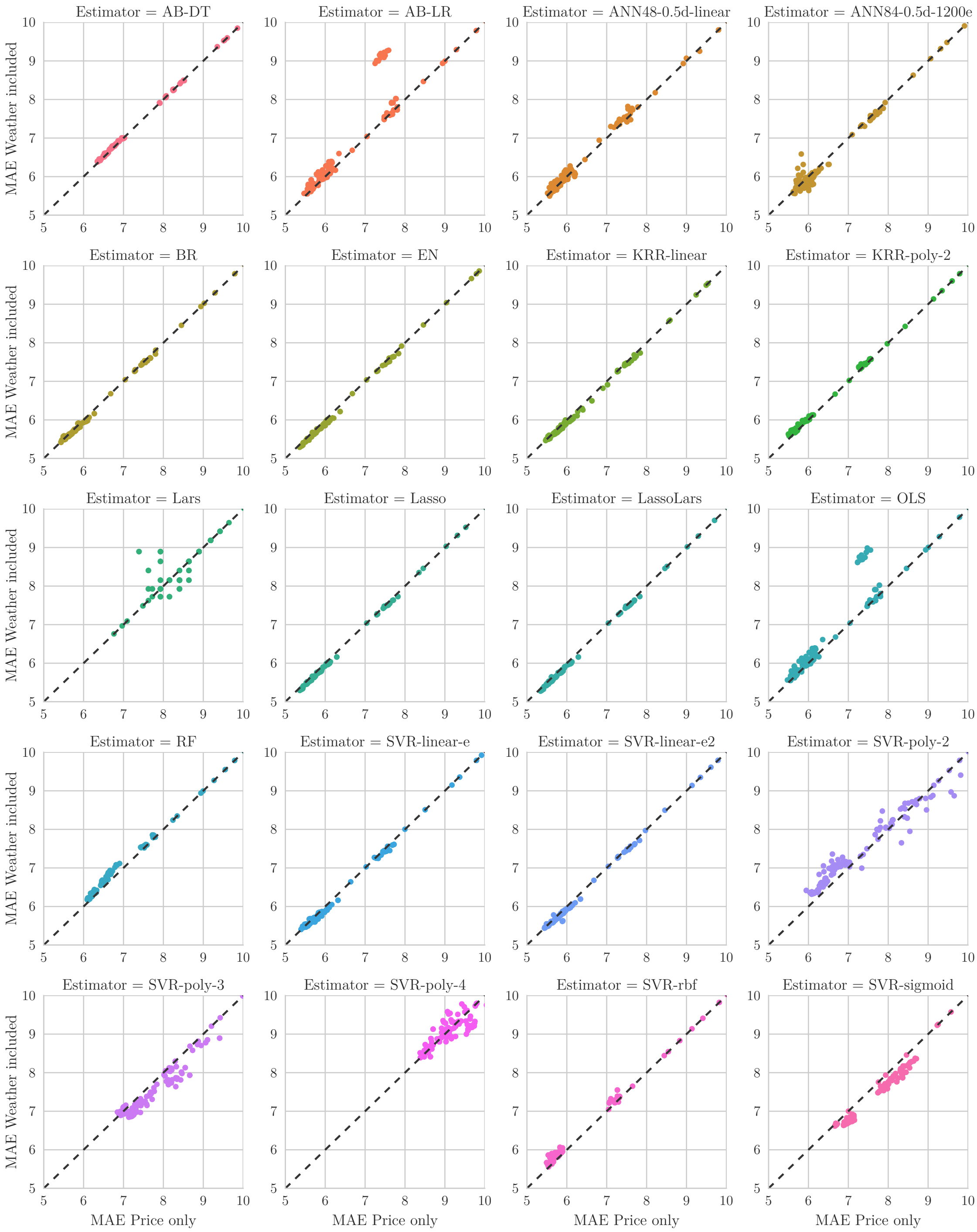
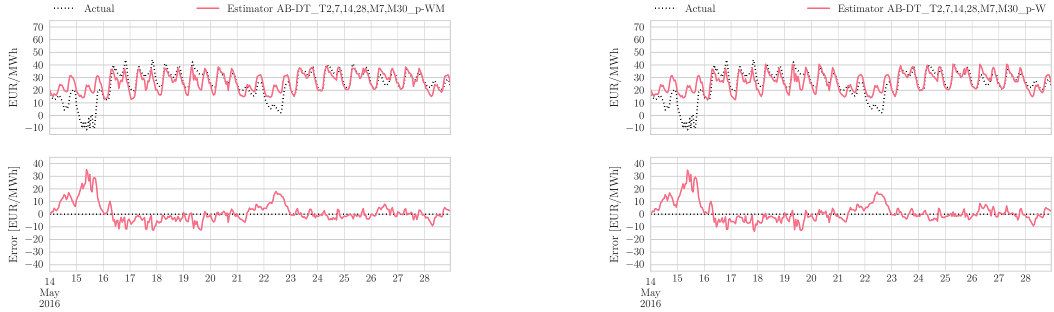


Figure D.2: The impact of weather data on MAE. Each point represent a pair of estimators A and B that differ only by the inclusion of weather data. The x axis shows the MAE of the estimator without the weather data and the y axis hows the MAE of the estimator with the weather data. The closer the estimators are to the dashed line, the smaller influence have the weather data on the estimator.

Appendix E

Examples of prediction

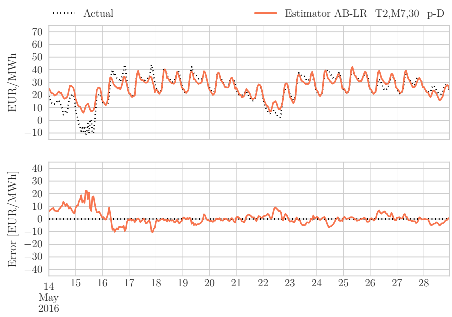
This appendix shows the examples of forecasts on the test data (out-of-sample) of best estimators in each estimator class for both RMSE and MAE measures.



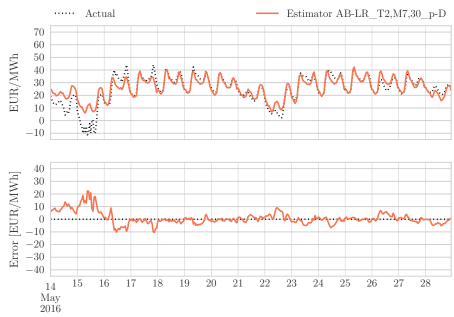
a by MAE

b by RMSE

Figure E.1: The best estimator of the `AB-DT` class.

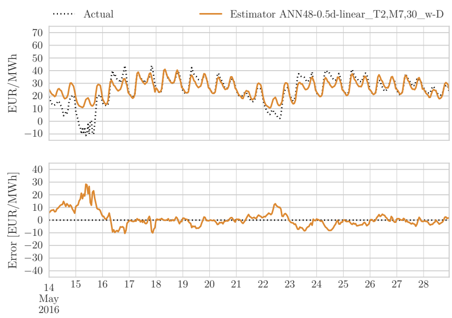


a by MAE

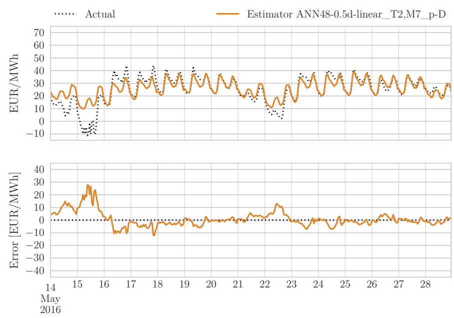


b by RMSE

Figure E.2: The best estimator of the `AB-LR` class.

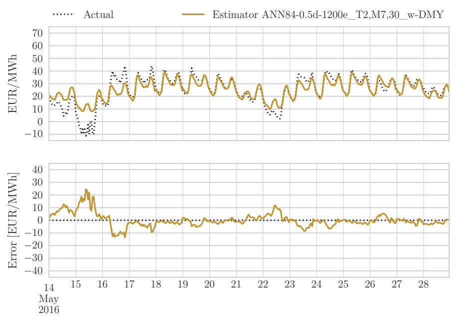


a by MAE

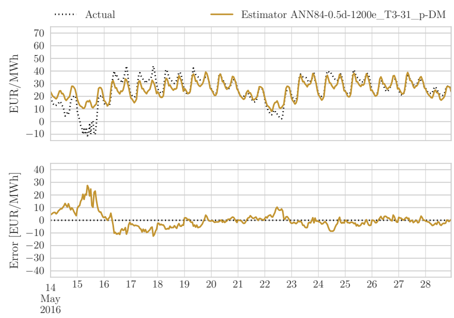


b by RMSE

Figure E.3: The best estimator of the `ANN48-0.5d-linear` class.

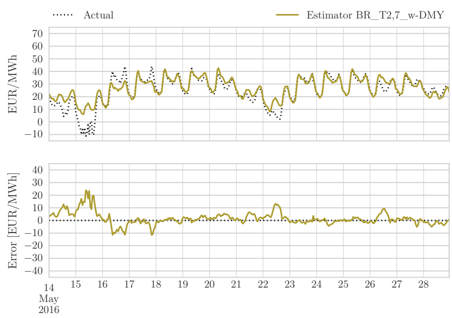


a by MAE

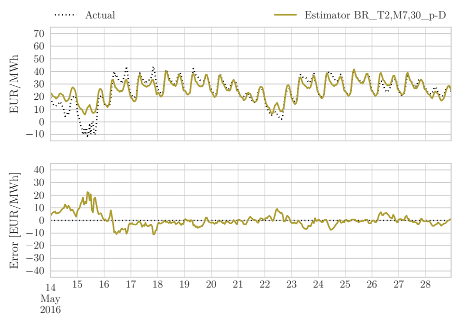


b by RMSE

Figure E.4: The best estimator of the `ANN84-0.5d-1200e` class.

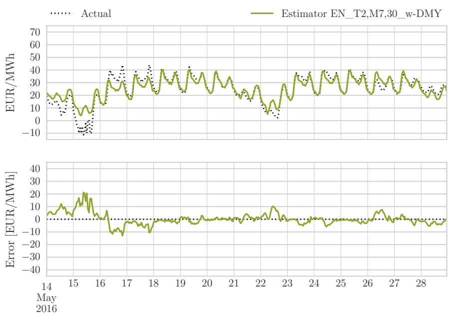


a by MAE

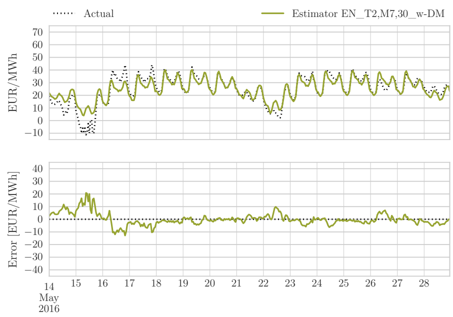


b by RMSE

Figure E.5: The best estimator of the `BR` class.

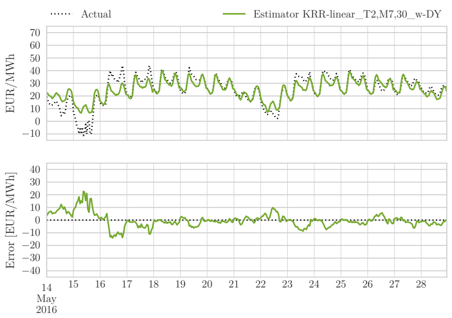


a by MAE

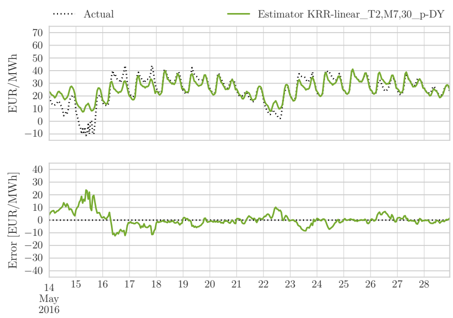


b by RMSE

Figure E.6: The best estimator of the `EN` class.

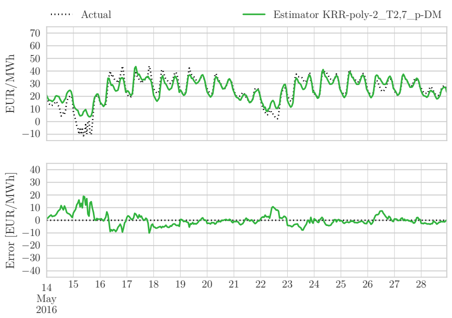


a by MAE

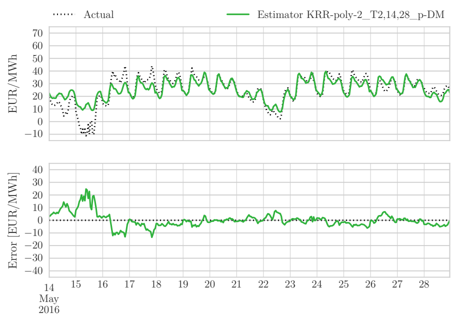


b by RMSE

Figure E.7: The best estimator of the `KRR-linear` class.

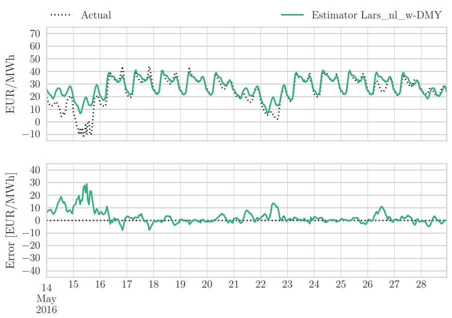


a by MAE

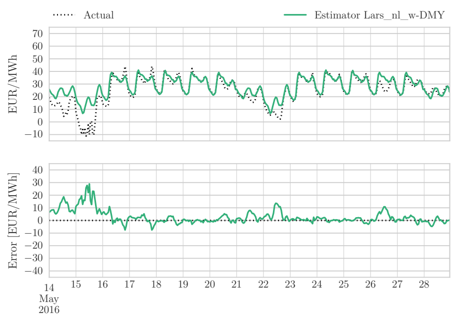


b by RMSE

Figure E.8: The best estimator of the `KRR-poly-2` class.

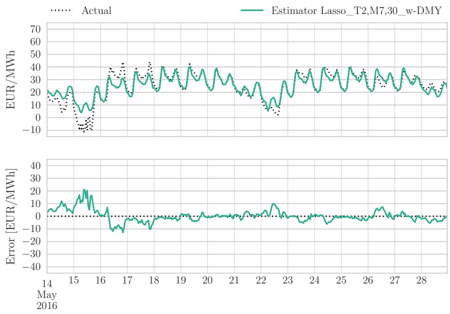


a by MAE

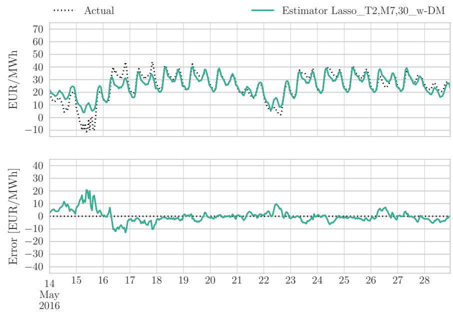


b by RMSE

Figure E.9: The best estimator of the `Lars` class.

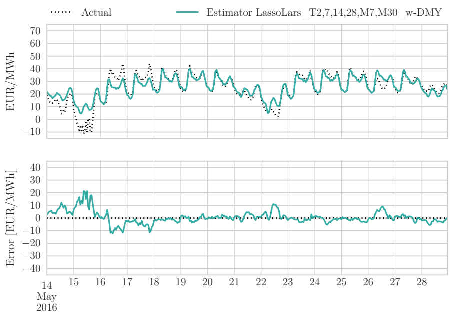


a by MAE

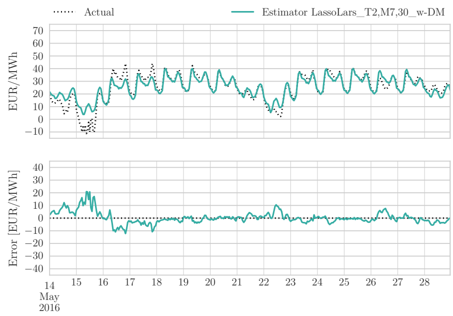


b by RMSE

Figure E.10: The best estimator of the `Lasso` class.

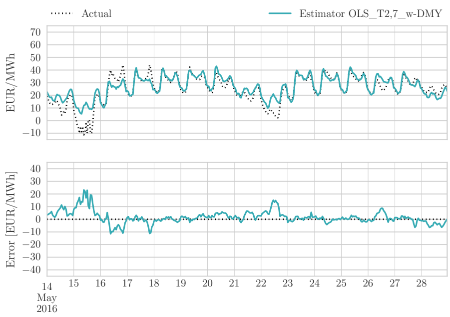


a by MAE

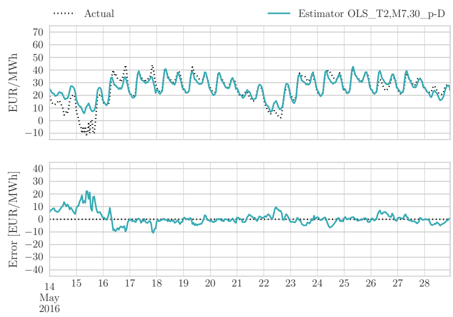


b by RMSE

Figure E.11: The best estimator of the `LassoLars` class.

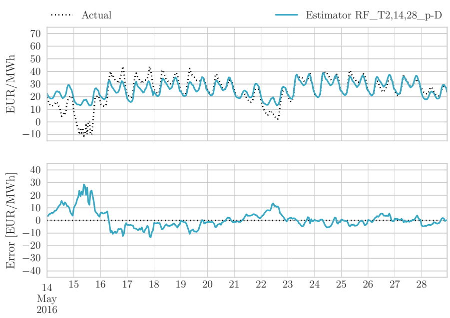


a by MAE

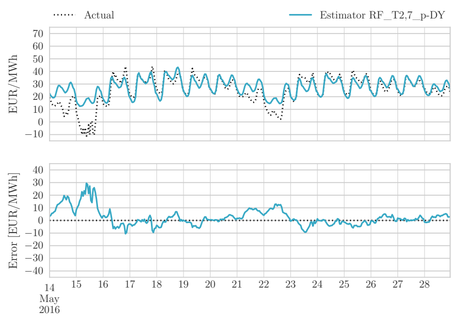


b by RMSE

Figure E.12: The best estimator of the `OLS` class.

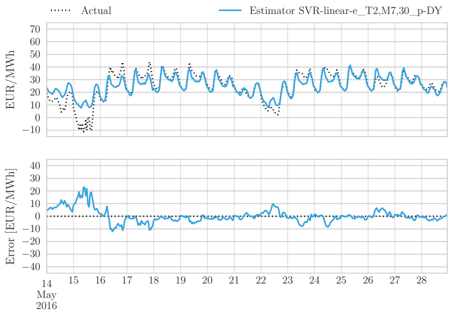


a by MAE

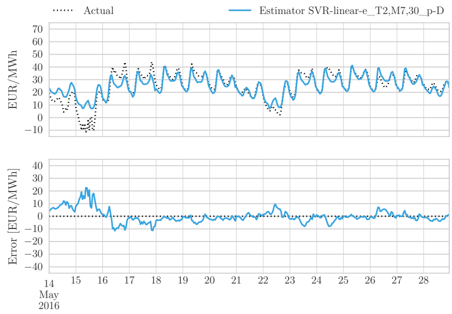


b by RMSE

Figure E.13: The best estimator of the `RF` class.

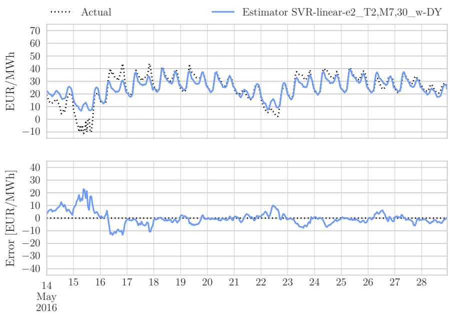


a by MAE

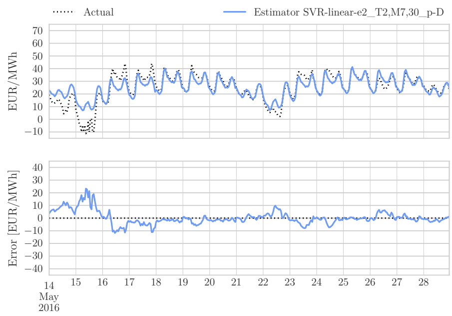


b by RMSE

Figure E.14: The best estimator of the `SVR-linear-e` class.

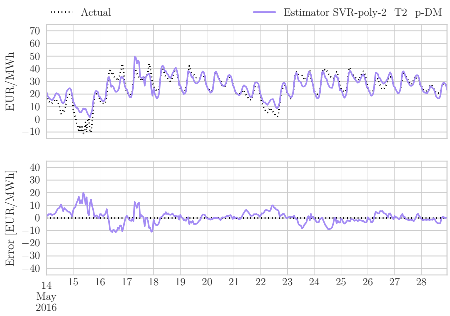


a by MAE

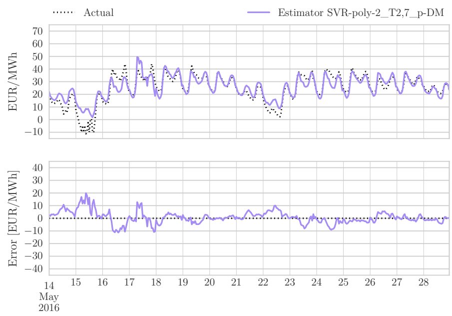


b by RMSE

Figure E.15: The best estimator of the `SVR-linear-e2` class.

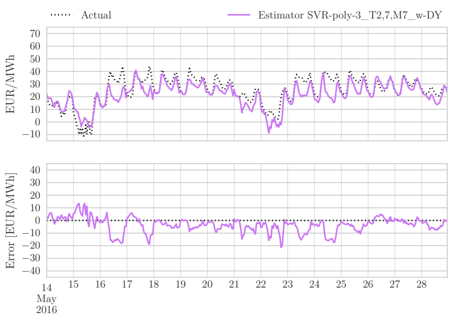


a by MAE

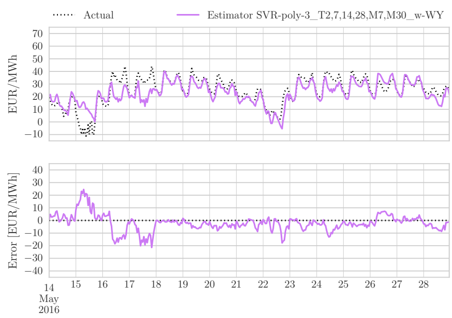


b by RMSE

Figure E.16: The best estimator of the `SVR-poly-2` class.

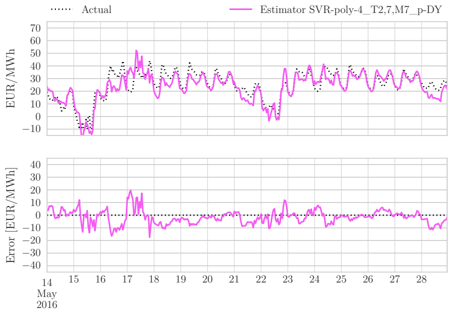


a by MAE

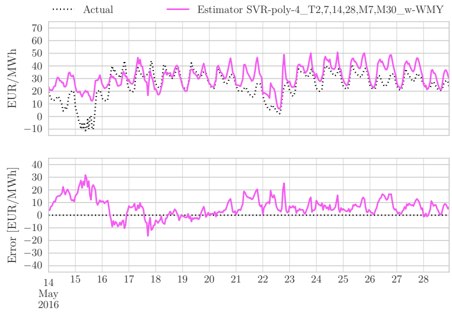


b by RMSE

Figure E.17: The best estimator of the `SVR-poly-3` class.

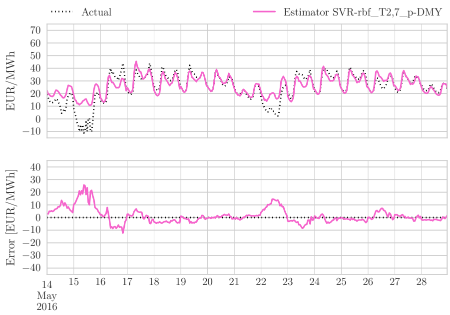


a by MAE

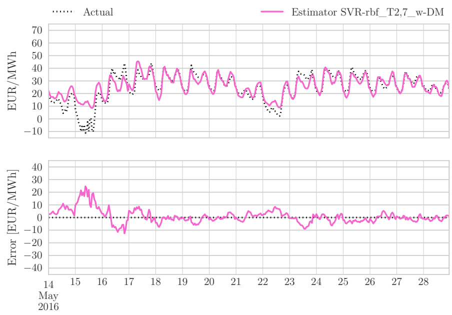


b by RMSE

Figure E.18: The best estimator of the `SVR-poly-4` class.

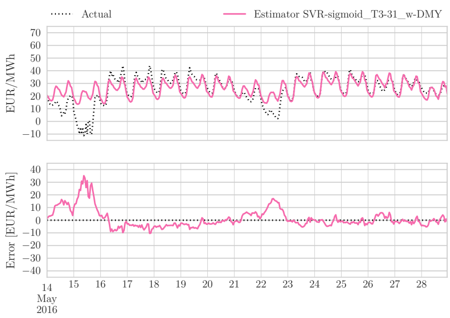


a by MAE

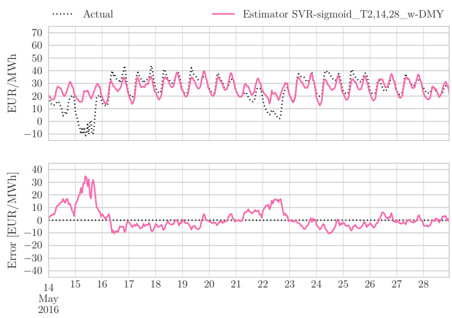


b by RMSE

Figure E.19: The best estimator of the `SVR-rbf` class.



a by MAE



b by RMSE

Figure E.20: The best estimator of the `SVR-sigmoid` class.