UNIVERSITY OF BERGAMO

School of Doctoral Studies

Doctoral Degree in Analytics for Economics and Business (AEB)

XXIX Cycle

SSD: Operations Research

OPTIMIZATION AND FORECASTING MODELS FOR ELECTRICITY MARKET AND RENEWABLE ENERGIES

Advisor:
Chiar.ma Prof.ssa Maria Teresa Vespucci

Doctoral Thesis

Federica DAVO'

Student ID 1031612

Academic year 2015/2016

Contents

Introduction **1**

1 Modelling framework **21**

1.1 Operations Research 21

1.1.1 Linear Programming (LP) 22

1.1.2 Integer Linear Programming (ILP) 25

1.2 Forecasting Models 28

1.2.1 Principal Component Analysis 28

1.2.2 Linear Regression 30

1.2.3 Neural Network 31

1.2.4 Analog Ensemble 32

1.2.5 Support Vector Regression 34

References 41

2 Post-processing Techniques and Principal Component Analysis for Regional Wind Power and Solar Irradiance Forecasting **47**

2.1 Introduction 47

2.2 Observational and meteorological data description 49

2.3 Prediction methods 52

2.4 Forecast evaluation 57

2.5 Results 61

2.6 Conclusions 69

References 71

3 Forecasting Italian electricity market prices using a Neural Network and a Support Vector Regression 77

3.1 Introduction 77

3.2 Data Description 79

3.3 Prediction Methods 84

3.3.1 Neural Network (NN) 84

3.3.2 Support Vector Regression (SVR) 85

3.3.3 Reference Methods 86

3.4 Results 87

3.4.1 Forecast evaluation 87

3.4.2 PUN results 88

3.4.3 Zonal Prices results 89

3.5 Conclusion 93

References 95

Introduction

The electricity world is in transition. In the last years, different factors are playing an important role in its development: the coupling of European electricity markets, the development of renewable energies combined with their large-scale integration, the research of the best investment strategy, are only some examples. Producers, consumers, market and network operators are all interested in having as much accurate information as possible. This is not easy to obtain, since the power system is subject to difficulties arising from different factors, as the variability of the demand, the complexities related to energy storage and the technical limits over the networks. These difficulties bring cogent constraints as the request of an instantaneous and continuous trade-off balance between the quantity of energy injected in and extracted from the network.

There are many examples of how the market participants can and want to use these informations. In the electricity market, for example, producers try to select the best strategies to maximize their profits. Since the producers interact in the market through different types of supply orders, these have become the main strategic instrument that power producers can use to manipulate the market. The solution of the market, however, is not only determined by these orders, and an inadequate selection of the supply curve may imply that the power market exchanges do not dispatch a producer. Therefore, to select an adequate offer strategy, each power producer must predict, as accurately as possible, the impact of its offers on the market outcomes. This motivates the development of different mathematical tools which allow market participants to find the optimal strategy or operators to better analyse the market behaviour.

Another example could be the power predictions that can be useful to renewable energy producers, but also to Transmission System Operators (TSO) or market operators. In particular, TSO can benefit from such predictions to satisfy different purposes (e.g.,

overloads management or reserve estimation). The energy market operators need accurate predictions to reduce penalties, which are usually proportional to the unbalances, defined as the difference between predicted and observed power. Some individual electricity producers have now penalties when they produce less than predicted or have to manage storage in batteries to ensure a smoother export to the grid as part of their contract. The considerations above motivate the definition of an adequate modelling tool that allows to forecast the power production in a short-term and on small or large areas.

This thesis presents different optimization and forecasting models, with the focus on energy markets and renewable energy sources. Most of the models have been done in collaboration with RSE - Ricerca Sistema Energetico, a company that has financed the Ph.D. scholarship, for conducting research with concerning "Optimization and forecasting models for electricity market and renewable energies". The analysis has been done principally from two different points of view, in order to better understand the market behaviour through different tools and at different levels. Initially, the Italian electricity day-ahead market has been studied. The other analysis approach is related to some forecasts useful for market operators and participants, in particular it is related to electricity price and power production forecasts. These tools aim the market operators to decide in advance different operations that can be done on the market.

This chapter introduces the thesis work reported in this dissertation. First, we motivate the thesis work, state its aim and describe the approach used to achieve the desired objectives. Then, a literature review of the topics pertaining to this dissertation is provided. Finally, the main objectives of the thesis are described together with a description of the chapters.

Market clearing problem

In the last years the coupling of the day-ahead European electricity markets has been subject to a steadily growing interest. In particular, the formulation of an exact model for solving market coupling is still considered a hot topic. In this respect Price Coupling

of Regions (PCR) [1] is an initiative of seven power exchanges - APX-ENDEX, Belpex, EPEX SPOT, GME, Nord Pool Spot, OMIE and OTE - that represent the electricity markets of Austria, Belgium, Denmark, Estonia, Finland, France, Germany, Italy, Latvia, Lithuania, Luxemburg, Norway, Netherlands, Poland, Portugal, Czech Republic, Spain, Sweden, Switzerland and United Kingdom. The initiative has been launched in 2009 and it is focused on the delivery of a common European single price coupling solution. The algorithm developed by PCR is named EUPHEMIA [2]. EUPHEMIA is used to calculate energy allocation and electricity prices across Europe, maximizing the overall welfare. It is the acronym of Pan-European Hybrid Electricity Market Integration Algorithm and it is developed to solve the problem associated to the European day-ahead market.

The European market is divided into several zones, and each area is expected to be driven by different prices. When all the electricity traded at the same location and at the same time period is exchanged at the same price then this price is referred to as linear. These prices are computed as optimal dual variables (shadow prices) of the demand-supply matching program [3][4]. With the assumption of non-convex constraints, that can hold for some electricity production [5], uniform, linear prices can not exist [6] or can not be defined as shadow prices of a demand-supply matching problem. This is the case of the European day-ahead market, with a variety of order types, as block orders, complex orders subject to a Minimum Income Condition (MIC) or orders subject to a Uniforme Purchase Price, as the Italian Prezzo Unico Nazionale (PUN) [2][7][8].

The complexities brought along by these orders are mainly related to their acceptance. A market clearing model could accept a block order even if it does not satisfy some price's definitions since in this case the welfare would be higher. In the PCR such orders are called paradoxically accepted and they are not allowed. Another complexity is given by Italian orders: supply orders are driven by the zonal prices, while the majority of demand orders are driven by PUN, except for some types of demand orders (as pumping demand orders) that are subject to the zonal prices. Hourly zonal prices, in a zonal market, are obtained by the intersection point between the demand curve and the supply curve in the considered market area. The hourly PUN is defined as the price that equals the total incomes with the total revenues (cost recovery constraint). Since the two different price signals cannot be both found as shadow prices of the demand-supply matching problem,

one must devise an ad-hoc procedure to clear the market obtaining both PUN and zonal prices. Furthermore, the PUN constraint cannot be simply added as a post-processing result since the acceptance of a class of purchase bids is related to the value of the PUN at the optimal solution.

Currently used approaches for the Italian and European electricity market clearing problem usually consist in the solution of different problems, each of them specific for a particular complexity of the entire market problem. In particular, the state of the art solver, EUPHEMIA, provides a solution to the pan-European Day-Ahead market coupling problem by solving a sequence of interrelated sub-problems. Namely, a combinatorial master problem is defined to maximize the social welfare without providing zonal market prices. After solving this problem, the framework solves three interdependent sub-problems to determine prices, PUN and overall exchanged volumes. When prices are determined in one of the sub-problems, the framework ensures that no orders are paradoxically accepted. Should this not happen, then the integer solution of the master problem is rejected and a cut is included in the master problem to prevent such solution to be picked in a subsequent iteration. If there is no paradoxical acceptance of the aforementioned orders, the framework proceeds to include its other sub-problems. One of the sub-problems of EUPHEMIA solves PUN problem with a procedure based on an heuristic iterative algorithm. The overall solution is then computed considering all the solutions together, and all these computations must be done within the available computational time. If the solution is not found in this time, an approximation on the different problems is done. This means that an optimal or exact (i.e. not approximated) solution is difficult to obtain, also because these approaches work with different problems, or on different iterations.

To our knowledge, an exact model without iterations, post-processing or sub-problems does not exist, neither for the Italian nor for the European day-ahead electricity market. This does not mean that these markets are not solved correctly, but the advantages of an unique model are various, as:

- The computational time of a unique model could be lower than the computational time of a model constituted by sub-problems, iterations or post-processing.

- If the unique model is well formulated, it could be easier and immediate to find the

optimal solution.

Forecasts

On the electricity market there are different players (as consumers, producers, traders), with different characteristics and with specific strategic objectives. For each player it is extremely important monitoring and forecasting the market trend on different temporal horizons for many reasons: in the short-medium term, for example, to manage the risk, in the medium-long term to plan the investment. To compute the market clearing price, all the orders (i.e. the offered quantities and prices) must be known. Without this information, it is possible only to forecast the market price, imaging a possible offer strategy and a possible future scenario. To create the best market offer, to make the higher profit or the best strategy, it is extremely important to have the right forecasts of all the useful information. For example, having a good forecast can mean reducing the penalties that must be paid by the producers for the unbalancing, which are computed as the difference between the promised energy and the delivered one, as shown in [9]. To this aim it would be good to have a forecast of the power production but also of the market price of the zone in which the producer is located.

Another example could be how price forecasts from few hours to few months ahead have become of particular interest to power portfolio managers. A generator, utility company or large industrial consumer who can forecast the volatile wholesale price with a reasonable level of accuracy can adjust its bidding strategy or its own production or consumption schedule in order to reduce the risk or maximize the profits in day-ahead trading.

A lot of research is focused in the activity of developing tools and algorithms for price, power production and load forecasting. A variety of methods and ideas have been tried for these forecasts, with varying degrees of success. Unfortunately, it has been observed that forecasting errors are still high from risk management perspective, from producers points of view, or for network security. For this reason in collaboration with RSE, different models and tools have been investigated and the variables that most influence electricity market objects, as prices and renewable energies, have been analysed. In particular, firstly

we decided to test a power production forecasting approach for large areas. This is due to the fact that in recent years, the development of wind and solar energy combined with their large-scale integration is creating a growing interest for predictions of the overall power production over large areas. These predictions can be used for reserve estimation, unbalances managing or to predict the electricity market prices. Furthermore, the supply orders of renewable energies offered on the electricity market come from power production forecasts, and a wrong prediction can lead to high penalties.

Afterwards, the role of different methods and predictors for the electricity market prices has been tested. Such predictions are useful since, for example, participants in deregulated electricity market can use price forecasting to develop their bidding strategies to maximize the profit obtained by trading energy. In bilateral contracts the agreed price of buyers and sellers is based on market clearing price predictions. On the other hand, the market behaviour can be analysed by the market operators using accurate price forecasts.

Literature review

Energy markets

The reference description of the European electricity markets is published in the documents of algorithm EUPHEMIA [2], the algorithm developed by the PCR [1] to solve the problem associated to the European day-ahead market. In this document it is possible to find the full description of all the European market structure (included the Italian market) and how the clearing problem is solved.

Italian (and so European) market is non-convex due to its complex orders and constraints. An extensive literature exists on non-convex day-ahead electricity markets, in particular references [6][7][10][11][12][13] and [14] show how the presence of indivisibilities in real-word markets preclude obtaining competitive prices. These prices do not maximize surplus to market participants as shown by [5][15][16] and [17].

Different methods have been proposed to find prices in a non-standard way, as [6] that presents an innovative methodology to find prices based on a reformulation of the non-convex market clearing problem where the integer variables are fixed at their optimal

value. In [18] is proposed a pricing approach based on a decentralized formulation of the electricity market, which considers simultaneously the problems faced by the generators and the system operators.

European electricity market is based on a zonal market, where optimal dual variables of the primal problem provide equilibrium prices [3][4]. A way to formulate common European market requirements in a mathematical model is via the addition of dual and complementary constraints to the primal program [4][19].

Other authors tried to find an exact solution of the European market clearing problem without iterations, among all [12][13], but a complete formulation, to our knowledge, is not given.

The Italian day-ahead electricity market has been coupled with the other European markets since February 2015, by linking the Italian border with France, Austria through the PCR (it was already linked with Slovenia). Up to this date, the Italian day-ahead electricity market was cleared with an heuristic iterative algorithm called UPPO [20]. UPPO was based on a Linear Programming (LP) model of a zonal market in which only the accepted quantities referred to orders subject to zonal prices were considered as variables; the sum of accepted quantities of orders subject to PUN and PUN itself were treated as parameters. The values of these parameters were fixed or calculated at the beginning of each iteration by means of sensitivity analysis on the LP model, then the LP model was solved and if the obtained solution along with the other parameters cleared the equation defining PUN, it was considered as the market clearing solution. UPPO, even if heuristic in nature, was able to find the correct solution, but was not suitable for easily being integrate in an European Market Coupling algorithm; furthermore, it was not able to solve markets with block orders or other kinds of complex orders which could be introduced in the Italian market. Anyway, EUPHEMIA is partially based on UPPO algorithm.

Currently used approaches for the Italian electricity market clearing problem usually solve a master problem with all the complex orders with a Mixed Integer Linear Program (MILP) [21][22][23] ignoring the PUN definition and considering all the Italian demand orders as subject to the zonal prices and a consecutive sub-problem that reintroduces the PUN definition and the partition between demand orders subject to PUN and demand orders subject to zonal price, with a possible relax requirement of the constraints related

to PUN. If a solution of the PUN sub-problem does not exist, some operations (i.e. cuts) on the master problem must be done to find a global solution for the updated master problem [2]. This method adds complexities to the problem, and might not lead to an exact solution, in particular when an upper bound on the available computational time is set for the identification of the solution.

Forecasts

Since electricity is difficult to storage, it is very important to have a good forecast of the renewable energy produced. For this reason in the last years the number of authors interested in this field is growing, and it is possible to find a huge literature. Different reviews and state-of-the-art papers describe perfectly the current situation, as [24][25] and [26]. Many different methods have been proposed. Kalman Filter are used by many authors, as in [27][28]. In [29] the authors use ARMA models for short-terms wind forecasts, while a linear autoregressive model and an adaptive fuzzy logic based model is used in [30]. Adaptive linear models, adaptive fuzzy logic models and wavelet based models are the methods used in [31][32], while in [33] is used an analysis of principal components of wavelets derived from wind speed time series for a measure-correlate-predict technique.

Large-scale wind power and solar irradiance forecasts have been the subject of several studies [34][35]. The most straightforward method to predict the power generated over an entire area is to sum the individual forecasts of each power plant. To do that, the location and characteristics of each plant must be known. Alternative methods can be proposed, which only require historical time-series of power measurements produced by all the plants located within the area, as the Principal Component Analysis (PCA) technique. PCA [36][37][38] is widely used in multivariate statistics and allows reducing the dataset size, retaining only the most relevant information. PCA is applied on the correlation matrix of the original predicted variables over all the grid points inside the area. A similar methodology was implemented to evaluate the performance of an analogous model on day-ahead forecasting of wind power production over large European regions in [39].

Price forecasting has been the subject of several studies (see e.g. [40] and [41] for a review), and a lot of different methods have been proposed for price and load forecast-

ing, such as linear regressions [42], stochastic processes [43], ARMA models [44], and also weighted nearest neighbours techniques [45], quantile regressions [46][47] or hybrid correction methods [48].

For the type of problem and available data in this thesis Neural Network, Analog Ensemble and Support Vector Regression are used. Neural network is a deterministic method [49][50], it is possible to use any different configuration of the network, and in literature there are a lot of examples, as [51][52][53], or rules [54][55] for the wind and solar power forecasts as long as in [56][57][58] for price forecasting.

Analog Ensemble method was originally proposed in [59]. It is a probabilistic prediction, and it can be used by the day-ahead electricity markets for trading future energy production, obtaining higher results than those obtained by using deterministic forecasts alone, as shown in [9][60]. Indeed, by using probabilistic forecasting, it is possible to optimize the revenue for a producer in an economic model depending for example on the specific penalties for forecast errors valid in that market. The analog ensemble produces probabilistic power predictions from historical deterministic predictions and observations of the quantity to be predicted, as in [46][61][62][63][64].

Support Vector Regression is based on Support Vector Machine, a new method proposed by Vapnik [65][66]. It is used a lot for price forecasting, with good results, as in [58][67] and [68].

Thesis

The main topic of the thesis is related to the forecasting models in the energy sector, in particular those for wind and solar power forecasts and those for electricity prices forecasts. These topics are important in the energy sector since, for example TSO can use wind and solar power forecasts for overloads management or reserve estimation, or can use price forecasting to analyse the market behaviour.

The thesis is composed of two parts. The first part includes this introduction and the modeling framework reported in Chapter 1.

The second part of the thesis consists in two chapters related to forecasting models. The first of them, Chapter 2, shows the application of a statistical methods and of two

forecasting methods (Neural Network and Analog Ensemble) to forecast the wind and solar power produced on large areas. Neural Network, together with Support Vector Regression, are the methods used in Chapter 3, for price forecasting.

The chapters related to models and methods are organized as follows.

Chapter 2: Post-processing Techniques and Principal Component Analysis for Regional Wind Power and Solar Irradiance Forecasting

This work explores a Principal Component Analysis in combination with two post-processing techniques for the prediction of wind power produced over Sicily, and of solar irradiance produced over the Oklahoma Mesonet. For wind power, the study is conducted over a 2-year long period, with hourly data of the aggregated wind power output of the island. The 0-72 hour wind predictions are generated with the limited-area Regional Atmospheric Model System, with boundary conditions provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) deterministic forecast. For solar irradiance, we consider daily data of the aggregated solar radiation energy output (based on the Kaggle competition dataset) over an 8-year long period. Numerical Weather Prediction data for the contest come from the National Oceanic & Atmospheric Administration – Earth System Research Laboratory (NOAA/ESRL) Global Ensemble Forecast System (GEFS) Reforecast Version 2. The PCA is applied to reduce the datasets dimension. A Neural Network and an Analog Ensemble post-processing are then applied on the PCA output to obtain the final forecasts. The study shows that combining PCA with these post-processing techniques leads to better results when compared to the implementation without the PCA reduction.

Chapter 3: Forecasting Italian electricity market prices using a Neural Network and a Support Vector Regression

This work explores two different techniques (a Neural Network and a Support Vector Regression) for the prediction of the Italian day-ahead electricity market prices, the zonal prices and the uniform purchase price (Prezzo Unico Nazionale or PUN). The study is conducted over a 2-year long period, with hourly data of the prices to be predicted and

a large set of variables used as predictors (i.e. historical prices, forecast load, wind and solar power forecasts, expected plenty or shortage of hydroelectric production, net transfer capacity available at the interconnections and the gas prices). A Neural Network and a Support Vector Regression are applied on the different predictors to obtain the final forecasts. Different predictors' combinations are analysed to find the best forecast. We compare the NN and SVR to two less sophisticated post-processing methods, i.e. a linear regression and the persistency.

References

[1] PCR. *PCR public presentation*, 2016. http://www.mercatoelettrico.org/it/MenuBiblioteca/Documenti/20160727_PCRStandardPresentation_detailed_PMU.pdf.

[2] PCR Market coupling algorithm. *EUPHEMIA public description, Version 0.6*, 2 October 2013. http://www.apxgroup.com/wpcontent/uploads/Euphemia-public-description-Nov-20131.pdf.

[3] S. A. Gabriel, A. J. Conejo, J. D. Fuller, B. F. Hobbs, and C. Ruiz. *Complementary modelling in energy markets*. Springer Publishing Company, Incorporated, 2012.

[4] M. Van Vyve. *Linear prices for non-convex electricity markets: models and algorithms*. Technical report, Université catholique de Louvain, Center for Operational Research and Econometrics (CORE), 2011.

[5] J. M. Arroyo and A. J. Conejo. *Optimal response of a thermal unit to an electricity spot market*. IEEE Transactions on Power Systems, vol. 15, n. 3, pp. 1098-1104, 2000.

[6] R. P. O'Neill, P. M. Sotkiewicz, B. F. Hobbs, M. H. Rothkopf and W. R. Stewart. *Efficient market-clearing prices in markets with nonconvexities*. European Journal of Operational Research, vol. 164, n. 1, pp. 269-285, 2005.

[7] N. S. Maria. *Day-ahead electricity market, proposals to adapt complex conditions in omel*. Master's thesis, Escuela Tcninca Superior de Ingeniera (ICAI), Universidad Pontificia Comillas, 2010.

[8] GME. *Vademecum of the Italian power exchange*, 2010. https://www.mercato elettrico.org/en/MenuBiblioteca/Documenti/20091112VademecumofIpex.pdf.

[9] S. Alessandrini, F. Davò, S. Sperati, M. Benini and L. Delle Monache. *Comparison of the economic impact of different wind power forecast system for producers*. Advances in Science Research, vol. 11, pp. 49-53, 2014.

[10] H. E. Scarf. *Mathematical programming and economic theory*. Operations research, vol. 38, pp. 377-385, 1990.

[11] H. E. Scarf. *The allocation of resource in the presence of indivisibilities*. Journal of Economic Perspectives, vol. 8, pp. 82-88, 1999.

[12] M. Madani and M. Van Vyve. *Computationally efficient (MIP) formulation and algorithms for European day-ahead electricity market auctions*. European Journal of Operational Research, vol. 242, n. 2, pp. 580-593, 2015.

[13] M. Madani and M. Van Vyve. *Minimizing opportunity costs of paradoxically rejected block orders in European day-ahead electricity markets*. European Energy Market (EEM), 11th International conference on the, pp. 1-5, 2014.

[14] R. P. O'Neill, P. M. Sotkiewics and M. H. Rothkopf. *Equilibrium prices in power exchanges with non-convex bids*. Technical report, working paper, 2007.

[15] W. R. Stewart, B. F. Hobbs, M. H. Rothkopf, U. Helman and R. P. O'Neill. *Solving for equilibrium pries in security-constrained unit commitment auction market for electric power using mixed-integer programming*. Proc. INFORMS annual meeting, San Antonio, TX, 2000.

[16] M. Madrigal and V. H. Quintana. *Existence and determination of competitive equilibrium in unit commitment power pool auctions: price setting ad scheduling alternatives*. IEEE Transaction on Power System, vol. 16, n. 3, pp. 380-388, 2001.

[17] A. L. Motto and F. D. Galiana. *Equilibrium of auction markets with unit commitment: the need for augmented pricing*. IEEE Transaction on Power System, vol. 17, n. 3, pp. 798-805, 2002.

[18] W. W. Hogan and B. J. Ring. *On minimum-uplift pricing for electricity market*. Unpublished, 2007.

[19] A. Martin, J. C. Muller and S. Pokutta. *Strict linear prices in non-convex European day-ahead electricity markets*. Optimization methods and software, arXiv:1203.4177.

[20] T. Caramanis and Associates Inc. *UPPO auction module user manual, appendix A: the standard hourly auction problem*, 2004. http://www.mercatoelettrico.org/It/MenuBiblioteca/Documenti/20041206UniformPurchase.pdf.

[21] M. Madani and M. Van Vyve. *A MIP framework for non-convex uniform price day-ahead electricity auctions*, 2015.

[22] A. L. Motto, F. D. Galiana, A. J. Conejo and J. M. Arroyo. *Network constrained multi-periods auction for a pool-based electricity market*. IEEE Transactions on Power Systems, vol. 17, n. 3, pp. 646-653, 2002.

[23] C. Ruiz. *Pricing non-convexities in an electricity pool*. IEEE Transactions on Power Systems, vol. 27, n. 3, pp. 1334-1342, 2012.

[24] G. Giebel, R. Brownsword, G. Kariniotakis, M. Denhard and C. Draxl. *The state-of-the-art in short-term prediction of wind power: a literature overview*. ANEMOS plus, 2011.

[25] S. Sperati, S. Alessandrini, P. Pinson and G. Karioniotakis. *The "weather intelligence for renewable energies" benchmarking exercises on short-term forecasting in wind and solar power generation*. Energies, vol. 8(9), pp. 9594-9619, 2015.

[26] X. Wang, P. Guoc and X. Huanga. *A review of wind power forecasting models*. Science direct, Energy Procedia, vol. 12, pp. 770-778, 2011.

[27] E. A. Bossanyi. *Short-term wind prediction using Kalman filter*. Wind Engineering, vol. 9, n. 1, pp. 1-8, 1985.

[28] H. Vihriälä, P. Ridanpää, R. Perälä, and L. Söderlund. *Control of a variable speed wind turbine with feed-forward of aerodynamic torque*. Proceedings of the European Wind Energy Conference, Nice, France,pp. 881-884. 1999.

[29] C. Tantareanu. *Wind Prediction in Short Term: A first step for a better wind turbine control.* Nordvestjysk Folkecenter for Vedvarende Energi, 1992.

[30] A. G. Dutton, G. Kariniotakis, J. A. Halliday, and E. Nogaret. *Load and wind power forecasting methods for the optimal management of isolated power systems with high wind penetration*. Wind Engineering, vol. 23, n. 2, pp. 69-87, 1999.

[31] G. Kariniotakis, E. Nogaret, and G. Stavrakis. *Advanced short-term forecasting of wind power production*. Proceedings of the European Wind Energy Conference held in Dublin, Ireland, pp. 751-754, 1997.

[32] G. Kariniotakis, E. Nogaret, A. G. Dutton, J. A. Halliday, and A. Androutsos. *Advanced wind power and load forecasting methods for the optimal management of isolated power systems*. Proceedings of the European Wind Energy Conference, Nice, France, pp. 1082-1085, 1999.

[33] H. Fukuda, S. Tamaki, M. Nakamura, H. Nagai, F. Shijo, S. Asat and K. Onaga. *Development of a wind velocity prediction method based on a data-mining type autoregressive model*. Proceedings of the European Wind Energy Conference, Copenhagen, Denmark, pp. 741-744, 2001.

[34] R. Girard, K. Laquaine and G. Karioniotakis. *Assessment of wind power predictability as a decision factor in the investment phase of wind farms*. Applied Energy, vol. 101, pp. 609-617, 2013.

[35] P. J. Trombe, P. Pinson and H. Madsen. *A general probabilistic forecasting framework for offshore wind power fluctuations*. Energies, vol. 5, pp. 621-657, 2012.

[36] I. T. Joliffe. *Principal component analysis*. Springer, New York, 1986.

[37] S. Wold, K. Esbensen and P. Geladi. *Principal component analysis*. Chemometrics and Intelligent Laboratory Systems, vol. 2, pp. 37-52, 1987.

[38] H. Abdi and L. J. Williams. *Principal component analysis*. vol. 2, pp. 433-459, 2010.

[39] M. L. Martin, F. Valero, A. Pascual, J. Sanz and L. Frias. *Analysis of wind power production by means of an analog model*. Atmospheric Research, vol. 143, pp. 238–249, 2014.

[40] S. K. Aggarwal, L. M. Saini and A. Kumar. *Electricity price forecasting in deregulated markets: a review and evaluation*. International Journal of Electrical Power and Energy System, vol. 31, n. 1, pp. 13-22, 2009.

[41] R. Weron. *Electricity price forecasting: a review of the state-of-the-art with a look into the future*. International Journal of Forecasting, vol. 30, n. 4, pp. 1030-1081, 2014.

[42] A. D. Papalexopoulos and T. C. Hesterberg. *A regression-based approach to short-term system load forecasting*. IEEE Transaction on Power System, vol. 5, n. 4, pp. 1535-1547, 1990.

[43] A. Kian and A. Keyhani. *Stochastic price modelling of electricity in deregulated energy markets*. Proceedings of the 34th annual Hawaii International Conference on System Science, pp. 1-7, 2001.

[44] J. C. Cuaresma, J. Hlouskova, S. Kossmeier and M. Obersteiner. *Forecasting electricity spot-prices using linear univariate time-series models*. Applied Energy, vol. 77, n. 1, pp. 87-106, 2004.

[45] A. Troncoso Lora, J. M. Riquelme Santos, A. Gomez Exposito and J. M. Martynez Ramos. *Electricity market price forecasting based on weighted nearest neighbours techniques*. IEEE Transaction on Power System, vol. 22, n. 3, pp. 1294-1301, 2007.

[46] H. A. Nielsen, H. Madsen and T. S. Nielsen. *Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts*. Wind Energy, vol. 9, pp. 95-108, 2006.

[47] K. Maciejowska, J. Nowotarski and R. Weron. *Probabilistic forecasting of electricity spot prices using factor quantile regression averaging*. International Journal of Forecasting, vol. 32, n. 3, pp. 957-965, 2016.

[48] T. Senjyu, P. Mandal, K. Uezato and T. Funabashi. *Next day load curve forecasting using hybrid correction method*. IEEE Transanction on Power System, vol. 20, n. 1, pp. 102-109, 2005.

[49] F. Cassola and M. Burlando. *Wind speed and wind energy forecast through Kalman filtering of numerical weather prediction model output*. Applied Energy, vol. 99, pp. 154-166, 2012.

[50] P. Pinson, H. A. Nielsen, H. Madsen and T. S. Nielsen. *Local linear regression with adaptive orthogonal fitting for the wind power application*. Statistics and Computing, vol. 18, pp. 59-71, 2008. http://link.springer.com/article/10.1007%2Fs11222-007-9038-7.

[51] U. E. Cali, B. Lange, J. Dobschinski, M. Kurt, C. Moehrlen and B. Ernst. *Artificial neural network based wind power forecasting using a multi-model approach*. 7th International Workshop on Large Scale Integration of Wind Power and on Transmission Networks for Offshore Wind Farms, 2008, Madrid.

[52] B. D. Ripley. *Pattern recognition and neural networks*. Cambridge University Press, 1996.

[53] L. Von Bremen. *Combination of deterministic and probabilistic meteorological models to enhance wind farm forecast*. Journal of Physics: Conference Series, vol. 75, 2007.

[54] Z. Boger and H. Guterman. *Knowledge extraction from artificial neural network models*. IEEE Systems, Man, and Cybernetics Conference, 1997.

[55] A. Blum. *Neural networks in C++*. Wiley, 1992.

[56] P. Mandal, T. Senjyu and T. Funabashi. *Neural networks approach to forecast several hour ahead electricity prices and loads in deregulated market*. Energy Conversion Manage, vol. 47, pp. 2128–2142, 2006.

[57] N. Amjady. *Day-ahead price forecasting of electricity markets by a new fuzzy neural network*. IEEE Transaction on Power System, vol. 21, n. 2, pp. 887-896, 2006.

[58] B. Chen, M. Chang and C. Lin. *Load forecasting using support vector machines: a study on EUNITE competition 2001*. IEEE Transaction on Power System, vol. 19, n. 4, pp. 1821-1830, 2001.

[59] S. Alessandrini, L. Delle Monache, S. Sperati and J. N. Nissen. *A novel application of an analog ensemble for short-term wind power forecasting*. Renewable Energy, vol. 76, pp. 768-781, 2015.

[60] M. S. Roulston, D. T. Kaplan, J. Hardenberg and L. A. Smith. *Using medium-range weather forecasts to improve the value of wind power production*. Renewable Energy, vol. 28, pp. 585-602, 2002. http://people.maths.ox.ac.uk/lenny/wind2003_roulston.pdf

[61] J. B. Bremnes. *A comparison of a few statistical models for making quantile wind power forecasts*. Wind Energy, vol. 9(1-2), pp. 3-11, 2006. http://onlinelibrary.wiley.com/doi/10.1002/we.182/full

[62] J. K. Møller, H. A. Nielsen and H. Madsen. *Time-adaptive quantile regression*. Computational Statistics & Data Analysis, vol. 52, pp. 1292–1303, 2006.

[63] J. Wang, A. Botterud, R. Bessa, H. Keko, L. Carvalho, D. Issicaba, J. Sumali and V. Miranda. *Wind power forecasting uncertainty an unit commitment*. Applied Energy, vol. 88, pp. 1014-1023, 2011.

[64] P. Pinson. *Estimation of the uncertainty in wind power forecasting*. Ph.D. dissertation, Ecole des Mines de Paris, 2006.

[65] V. N. Vapnik. *The nature of statistical learning theory*. Springer, NY, 1995.

[66] V. N. Vapnik. *Statistical learning theory*. Wiley, NY, 1998.

[67] C. J. C. Burges. *A tutorial on support vector machines for pattern recognition*. Data Mining and Knowledge Discovery, vol. 2, pp. 121-167, 1998.

[68] C. Gao, E. Bompard, R. Napoli and H. Cheng. *Price forecast in the competitive electricity market by support vector machine*. Physica A: Statistical Mechanics and its application, vol. 382, n. 1, pp. 98-113, 2007.

Chapter 1

Modelling framework

1.1 Operations Research

Operations research can be defined as a science which uses quantitative methods to support decision making in order to take the best decision [1][2]. Mathematical programming is the study of problems in which a real function must be maximized or minimized choosing the values of real or integer variables from an allowed set. In what follows, a short introduction of the mathematical programming models is given. It is not a complete overview, but references to literature are given for each model. An introduction to the field of operations research can be found in [3].

An optimization problem can be represented in the following way:

Given:	a function $f : A \to \mathbb{R}$
Find:	an element $x_0 \in A$ s.t. $f(x_0) \leq f(x) \forall x \in A$ in case of minimization
	an element $x_0 \in A$ s.t. $f(x_0) \geq f(x) \forall x \in A$ in case of maximization

The subset $A \subseteq \mathbb{R}^n$, called *search space* or *choice set*, is specified by a set of constraints, equalities or inequalities that the elements of A have to satisfy. The elements of A are called *feasible solutions*. The function f is called in general *objective function*, otherwise *loss function* or *cost function* when must be minimized [4], *utility function* or *fitness function* when must be maximized. A feasible solution that minimizes (or maximizes) the objective function is called an *optimal solution*.

Operations research has different sub-fields, each of them called with a name that describes the structure of the domain or the proprieties of the objective function.

1.1.1 Linear Programming (LP)

Linear programming (LP) is a sub-field of operations research in which the problem is represented by linear relationships. More formally, linear programming is a technique for the optimization of a linear objective function, subject to linear equality and linear inequality constraints. The general LP problem can be formulated in the following way:

$$\max c^T x \tag{1.1}$$

$$\text{s.t. } Ax \leq b \tag{1.2}$$

$$x \geq 0 \tag{1.3}$$

The decision variable in the problem is given by x, while c^T, A and b are known parameters. Equation (1.1) is the objective function, equation (1.2) gives the constraint on the decision variable x. Equation (1.3) is the non-negativity constraint on x. Equations (1.2) and (1.3) define the feasible set for x.

A vector x for a linear programming problem is said to be *feasible* if it satisfies the corresponding constraints. A linear programming problem is said to be *feasible* if the constraint set is not empty; otherwise it is said to be *infeasible*. A feasible vector at which the objective function achieves the maximum (or minimum) value is called *optimal*. The feasible region is a convex polytopes, which is a set defined as the intersection of finitely many half spaces, each of which is defined by a linear inequality.

For example, let us suppose to have the following objective function

$$f(x) = 3x_1 + 4x_2$$

subject to the following constraints:

$$x_1 + 2x_2 \leq 14$$

$$3x_1 - x_2 \geq 0$$

$$x_1 - x_2 \leq 2$$

$$x_1, x_2 \geq 0$$

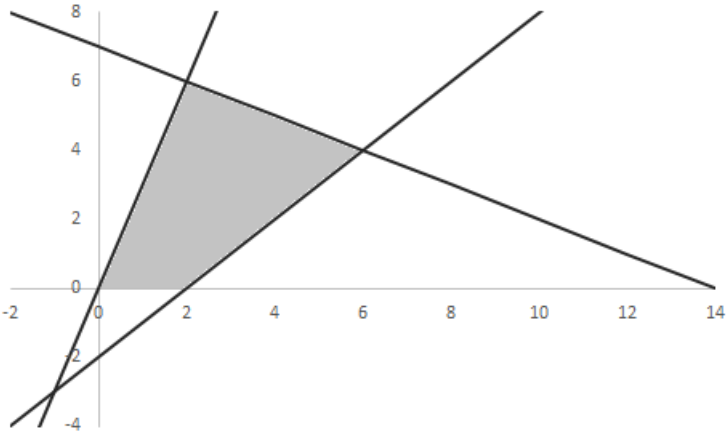


Figure 1.1: *Convex polytopes defined by the inequalities of the example.*

Its feasible region is shown in Figure 1.1.

The objective function of a LP problem is a real-valued affine function defined on this polyhedron. A linear programming algorithm finds a point in the polyhedron where this function has the smallest (or largest) value if such a point exists.

A linear programming problem is usually solved with the *simplex method*. Developed by George Dantzig in 1947 [5][6], it has proved to be a remarkably efficient method that is used routinely to solve huge problems on today's computers [3]. Except for its use on tiny problems, this method is always executed on a computer, and sophisticated software packages are widely available. A description of how to solve a linear programming problem through the simplex method can be found in [3][7].

In the general case in which the objective function, the constraints or both contain non linear terms the problem is called *non linear*. Solving a non-linear program is harder that solving a linear program, and the optimal solution it is not always guaranteed. Different methods and algorithms to solve a non linear program are shown in [8][9].

Dual problem

To every linear program (called *primal problem*) there is a *dual program* with which it is strictly connected. Given the standard form for the primal problem (1.1) - (1.3), its

dual problem has the following form:

$$\min y^T b \tag{1.4}$$

$$\text{s.t. } Ay \geq c \tag{1.5}$$

$$y \geq 0 \tag{1.6}$$

Thus, the dual problem uses exactly the same parameters as the primal problem, but in different locations. Let us summarize the relationships between the primal and dual problems.

Theorem 1.1 (Weak duality property)**.** If x is a feasible solution for the primal problem and y is a feasible solution for the dual problem, then

$$c^T x \leq y^T b.$$

Theorem 1.2 (Strong duality property)**.** If x^* is a optimal solution for the primal problem and y^* is a optimal solution for the dual problem, then

$$c^T x^* = y^{*T} b.$$

We call *duality gap* the difference between the primal and dual solutions, $c^T x^* - y^{*T} b$. The duality gap is zero if and only if strong duality holds [10].

Karush–Kuhn–Tucker conditions

The Karush–Kuhn–Tucker (KKT) conditions are first order necessary conditions for a solution in linear and non-linear programming to be optimal, provided that some regularity conditions are satisfied. Given the explicit form of a general problem:

$$\begin{aligned} &\min_{x \in \mathbb{R}^n} f(x) \\ \text{subject to } &h_i(x) \leq 0 \quad i = 1, \ldots, m \\ &l_j(x) = 0 \quad l = 1, \ldots, r \end{aligned}$$

and the correspondent dual problem

$$\begin{aligned} &\max_{u \in \mathbb{R}^m, v \in \mathbb{R}^r} g(u, v) \\ &\text{subject to } u_i \geq 0 \end{aligned}$$

The KKT conditions are:

$$0 = \partial f(x) + \sum_{i=1}^{m} u_i \cdot \partial h_i(x) + \sum_{j=1}^{r} v_j \cdot \partial l_j(x) \quad \text{stationarity}$$

$$u_i \cdot h_i(x) = 0 \quad \forall i \qquad \text{complementary slackness}$$

$$h_i(x) \leq 0, \quad l_j(x) = 0 \quad \forall i, j \qquad \text{primal feasibility}$$

$$u_i \geq 0 \quad \forall i \qquad \text{dual feasibility}$$

The following two proprieties hold [11][12]:

Necessity If x^* and u^*, v^* are primal and dual optimal solutions, and if the duality gap is zero, then x^*, u^* and v^* satisfy the KKT conditions.

Sufficiency If x^*, u^* and v^* satisfy the KKT conditions, then they are primal and dual optimal solutions.

Linearization Techniques

In KKT conditions there are products between primal and dual variables, as $u_i \cdot h_i(x) = 0$. If we want to solve a LP problem, these products cannot be inserted, otherwise the problem would become non-linear. Therefore they must be considered using linear constraints and this is possible using disjunctive constraints. Namely, for each constraint of the form $u_i \cdot h_i(x) = 0$ it is possible to add a binary variable y_i to the problem and use the following auxiliary constraints with a big $M \gg 1$ [13]:

$$u_i \leq M \cdot y_i$$

$$h_i(x) \leq M \cdot (1 - y_i)$$

1.1.2 Integer Linear Programming (ILP)

In a LP problem, if the variables are logical (with values 0 or 1) or integer, then the problem is called Integer Linear Programming (ILP). The general formulation is given by:

$$\max c^T x$$

$$\text{s.t. } Ax \leq b$$

$$x \in \mathbb{Z}_+^n$$

Figure 1.2 shows the geometric representation of such a problem, where the feasible region is given by the only points falling in the crossing of the squared area.

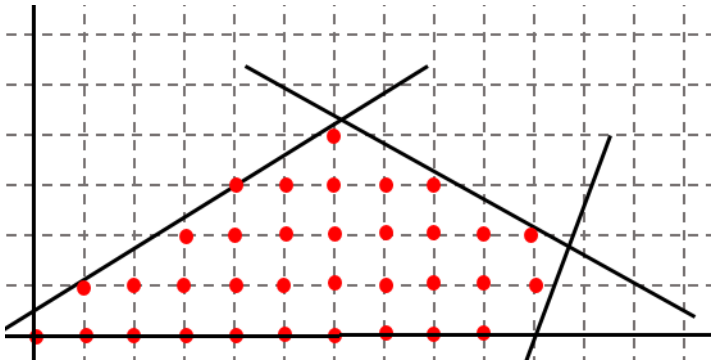


Figure 1.2: *The points of the feasible region of the ILP problem are marked in red.*

Note that the feasible region (consisting of a discrete set of points) is not a convex set anymore, as it was the case in linear programming. Consequently, the theory developed for LP cannot be directly applied to this class of problems.

Operation research includes also the class of Mixed Integer Linear Programming (MILP) problems, that are LP problems in which only a subset of variables has integer values. Here is an example of MILP problem:

$$\begin{aligned} &\max\ c^T x + d^T y \\ &\text{s.t. } Ax + Dy \leq b \\ &\qquad x \geq 0, y \in \mathbb{Z}^n_+ \end{aligned}$$

Branch and Bound

There are different algorithms to find the solution of an IP (or MILP) problem, one of the principals is the Branch and Bound that uses linear programming relaxation to find the optimal solution. This is the basic algorithm used by all commercial codes for solving MILP problem. Here, for simplicity of notation, is reported the presentation of IP. The idea is solving an IP in linear relaxation and finding an optimal solution z_0. If z_0 is integer the algorithm stops, otherwise, let us take one of the variables that has a fractional value, e.g. $x_1 = x_1^*$, where $x_1 \in \mathbb{Q} \setminus \mathbb{Z}$, and consider $\lfloor x_1^* \rfloor$ and $\lfloor x_1^* \rfloor + 1$. The initial IP problem is split now in two other sub-problems 1 and 2, that correspond respectively to the initial IP

problem plus constraint $x \leq \lfloor x_1^* \rfloor$ and to the initial IP problem plus constraint $x \geq \lfloor x_1^* \rfloor + 1$, as shown in Figure 1.3.

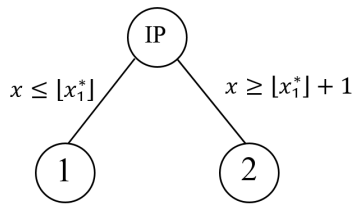


Figure 1.3: *Branching on the initial IP problem.*

Now, on problems 1 and 2 it is possible to solve a linear relaxation and restart with the algorithm. It is possible to proceed branching each node and pruning it if:

- The relaxation on the node gives an optimal solution that is integer.
- The relaxation on the node doesn't give any feasible solution.
- The relaxation on the node gives an optimal solution that is lower than an optimal integer solution obtain on another node.

Let us consider the general integer programming problem in the form

$$z_{IP} = \max\{c^T x : x \in S\} \qquad \text{where} \qquad S = \{x \in \mathbb{Z}_+^n : Ax \leq b\}$$

The general branch-and-bound algorithm is the following [14]:

Step 1 (Initialization): $\mathscr{L} = \{\text{IP}\}$, $S^0 = S$, $\overline{z}^0 = +\infty$ and $\underline{z}_{IP} = -\infty$.

Step 2 (Termination test): If $\mathscr{L} = \emptyset$, then the solution x^0 that yielded $\underline{z}_{IP} = cx^0$ is optimal.

Step 3 (Problem selection and relaxation): Select and delete a problem IP^i from $\mathscr{L}$. Solve its relaxation RP^i. Let z_R^i be the optimal value of the relaxation and let x_R^i be an optimal solution if one exists.

Step 4 (Pruning): a. If $z_R^i \leq \underline{z}_{IP}$, go to Step 2.

b. If $x_R^i \in S^i$, go to Step 5.

c. If $x_R^i \in S^i$ and $cx_R^i > \underline{z}_{IP}$, let $\underline{z}_{IP} = cx_R^i$. Delete from $\mathscr{L}$ all problems with $\overline{z}^i \leq \underline{z}_{IP}$. If $cx_R^i = \underline{z}_{IP}$, go to Step 2; otherwise go to Step 5.

Step 5 (Division): Let $(S^{ij})_{j=1}^{k}$ be a division of S^i. Add problems $\{\mathrm{IP}^{ij}\}_{j=1}^{k}$ to $\mathscr{L}$, where $\bar{z}^{ij} = z_R^i$ for $j = 1, \ldots, k$. Go to Step 2.

1.2 Forecasting Models

Forecasting means predict the future considering past and present data and analysis of trends. Many different methods have been proposed for forecasting framework. It is possible to take a classification of groups of models as follows [15][16][17]:

- *Game-theory* models, which simulate the operation of a system of heterogeneous agents interacting with each others.
- *Simulation* models, which describe the dynamics by modelling the impacts of physical factors.
- *Stochastic* models, which characterize the statistical proprieties over time.
- *Regression* models, based on the theorized relationship between a dependent variable and a number of independent variables that are known or can be estimated.
- *Artificial intelligence* models, which combine elements of learning, evolution and fuzziness to create approaches that are capable of adapting to complex dynamic system.

In this section all the forecasting methods used in the thesis will be shown. Firstly, the Principal Component Analysis (PCA) are introduced since they are used in Chapter 2 before using the forecasting methods.

1.2.1 Principal Component Analysis

A Principal Component Analysis (PCA) models the variation in a set of variables in terms of a smaller number of independent linear combinations (principal components) of those variables. PCA is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. Since patterns can be hard to

find in data of high dimension, PCA is a powerful tool for analysing data. Furthermore, with PCA is possible to reduce the data dimension, without loss of information.

Let us suppose to have two variables x_1 and x_2 as in [18], plotted in Figure 1.4, highly correlated with one another. It is possible to pass a vector through the long axis of the cloud of points and a second vector perpendicular to the first, with both vectors passing through the centroid of the data.

Now, it is possible to find the coordinates of all the data points relative to these two perpendicular vectors and replot the data, as shown in Figure 1.5.

In this new reference frame, the variance is greater along axis 1 than on axis 2. Furthermore, the spatial relationships of the points are unchanged; this process has merely rotated the data.

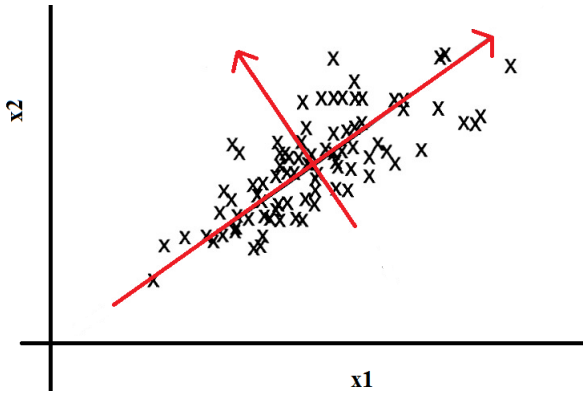


Figure 1.4: *Example of observable variables. The red arrows indicate the directions of the eigenvalues of the correlation matrix.*

Formally, given a data matrix X with p variables and n samples, the data are first centred on the means of each variable, i.e. the mean is subtracted from each of the data dimension. This will insure that the cloud of data is centred on the origin of the principal components, but does not affect the spatial relationships of the data nor the variances along the variables. The second step is compute the covariance matrix of data, and compute the eigenvalues and eigenvector of this matrix. The eigenvectors, also called *loadings*, are perpendicular to each others (as the red arrows in Figure 1.4) and provide information about the patterns in the data. To obtain the principal components, it is sufficient to multiply the data X by the matrix of eigenvectors.

Up to now, the original data have been transformed in principal components, each of

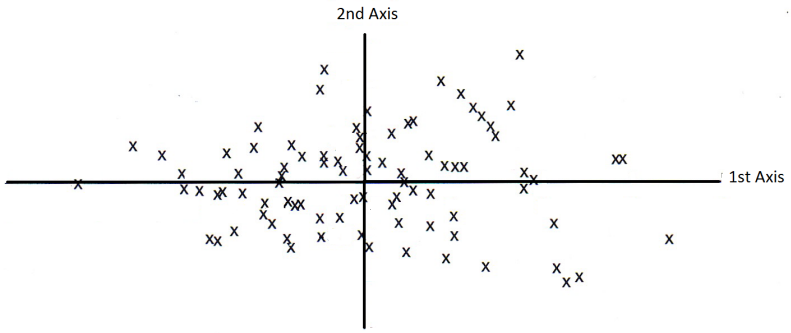


Figure 1.5: *Object transformation in the new Cartesian System.*

them non-correlated with the others, but the dimension is the same of the original data. To reduce the dimension from p to m, it is sufficient to take only the first m principal components.

In the literature three methods [19] have been introduced to choose the dimension p:

1. Cumulative percentage of total variation: only those components representing the $80\% - 90\%$ of the total variance are retained, i.e.,

$$\frac{\lambda_1 + \cdots \lambda_p}{\lambda_1 + \cdots \lambda_m} \approx 80\% - 90\% \tag{1.7}$$

   where λ_i is the i-th eigenvalues of the correlation matrix; the numerator represents the variance of the first p principal components, whereas the denominator represents the variance of all the main components.

2. The so-called Kaiser rule (proposed by Kaiser [20]) takes only those components having an eigenvalue greater than one.

3. Doing a "Scree Plot" where the number of principal components corresponds to the change of slope in the Scree Plot.

1.2.2 Linear Regression

Linear regression (LR) allows to compute the linear dependence between one or more independent variables and a dependent variable, following the formula:

$$y = \beta_0 + \sum_{i \in I} \beta_i \cdot x_i + u$$

where i varies between the predictors, y is the dependent variable, x_i are the independent variables, $\beta_0 + \sum_{i \in I} \beta_i \cdot x_i + u$ is the regression line, β_0 is the regression intercept, β_i is the angular coefficient of x_i and u is the statistical error.

1.2.3 Neural Network

A Neural Network (NN) is a system of interconnected group of nodes, called neurons since they remember the neurons of a human brain. They are able to link some input variables to one or more output variables, as shown in [21][22][23]. The neurons are connected by links and they interact with each other. The nodes can take input data and perform simple operations on them. The result of these operations is passed to other neurons. The output at each node is called activation or node value.

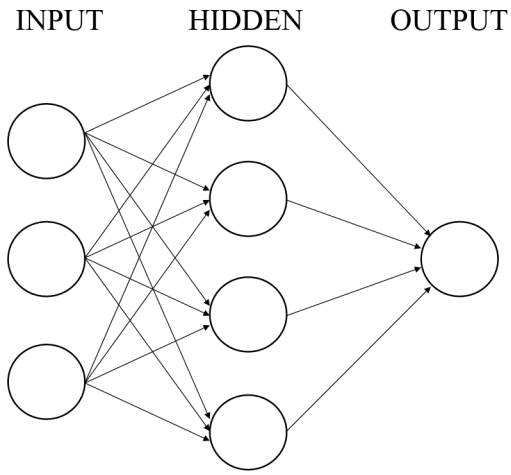


Figure 1.6: *Representation of a general Neural Network.*

To this end, each neuron assigns a weight to the input signal and with a learning process performed through different iterative cycles it learns how to associate the right output to each input.

Given the input $x_1, x_2, \ldots, x_n$ the output term is given by

$$h_W(x) = f(W^T x)$$

where $f : \mathbb{R} \to \mathbb{R}$ is called the activation function and W are the weights. One of the possible activation functions is given by the Hyperbolic Tangent $f(x) = tanh(ax^T + c)$ [24].

There are two Neural Network types: Feed-Forward and Feed-back. In the Feed-Forward NN the information flow is unidirectional. A unit sends information to other units from which it does not receive any information, this means that there are no feedback loops. An example of Feed-Forward NN is shown in Figure 1.6.

In the Feed-back NN loops are allowed, as shown in Figure 1.7. In this case, units can send information in both directions.

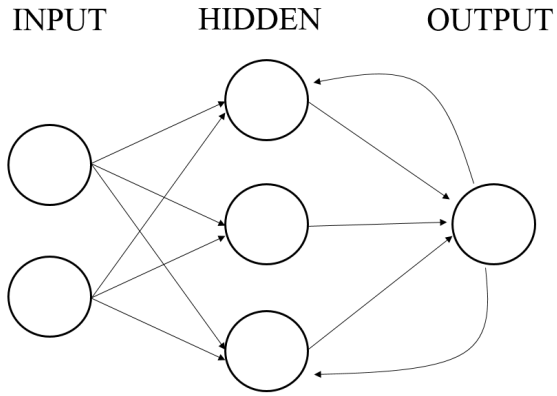


Figure 1.7: *Representation of a Feed-back Neural Network.*

NN are capable of learning and they need to be trained. There are several learning strategies, those used in this thesis is the Supervised Learning. This strategy needs a training period in which the NN learns how to associate to each input the correspondent output. This is done passing to the NN the input and the correspondent measured output in the training period.

1.2.4 Analog Ensemble

The Analog Ensemble (AnEn) method builds an ensemble forecast whose members are made by past observations (analogs) appropriately selected across a historical dataset of deterministic forecasts. An extended literature can be found for example in [25][26] and [27].

The AnEn method as proposed in [28] is a technique to generate an uncertainty forecast from a purely deterministic prediction. The uncertainty information is estimated using a set of M past verifying observations that correspond to the M past forecasts (analogs), which are most similar to a current deterministic forecast. Since this approach

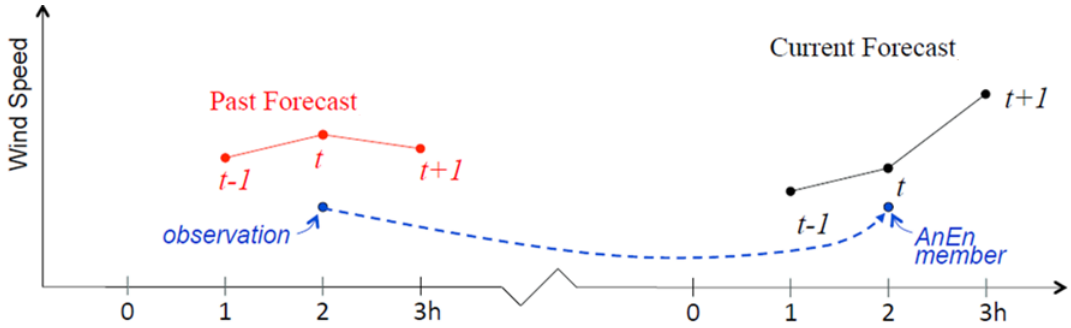


Figure 1.8: *AnEn is based on building an ensemble forecast whose members are past observation.*

directly uses the verifying observations as ensemble members, the AnEn method automatically accounts for observational errors in the verification. The metric used to estimate the degree of analogy between the current forecast and past predictions from a historical data set is defined as

$$|F_t, A_{\bar{t}}| = \sum_{i=1}^{N_v} \frac{w_i}{\sigma_{f_i}} \sqrt{\sum_{j=-\tau}^{j=\tau} \left(F_{i,t+j} - A_{i,\bar{t}+j}\right)^2} \tag{1.8}$$

where F_t is the current forecast at time t at a certain location, $A_{\bar{t}}$ is an analog forecast at time $\bar{t}$ before F_t was issued and at the same location, N_v and w_i are the number of physical variables and their weights, σ_{f_i} is the standard deviation of the time series of past forecasts of a given variable at the same location; τ is an integer equal to half width of the time window over which the metric is computed, and $A_{i,\bar{t}+j}$ and $F_{i,t+j}$ are the values of the analog and the forecast in the time window for a given variable.

The goal is to find past forecasts of the variables (chosen among those with the highest correlation with the quantity to be predicted) that were predicting similar values and temporal trend compared to the current forecast. The assumption is that if these forecasts are found, their errors will likely be similar to the error of the current forecast, which can be inferred from theirs. The main steps of the algorithm can be summarized as follows [25]:

1. Retrieve an historical dataset of predictions.

2. Retrieve the correspondent historical dataset of measurements.

3. Choose the physical variables from the model to be used as predictors in Equation (1.8) for the predictand variable.

4. For each lead time of the current forecast compute the distance (i.e., Equation (1.8)) from every past forecast issued at the same lead time.

5. For each lead time, rank all the past forecasts from the lowest distance and keep the n closest.

6. The concurrent n past measurements are the n members that constitute the current analog ensemble forecast for the lead time considered.

The AnEn provides probability density functions (PDF) from which the likelihood of future events can be estimated. The PDF of the power forecast is estimated using a set of n past verifying observations corresponding to the n best analogs (past model predictions) to a current deterministic model forecast. The verifying observation for each analog is thus a member of the AnEn.

1.2.5 Support Vector Regression

Support Vector Regression (SVR) is based on Support Vector Machine (SVM), a new and promising technique for data classification and regression, introduced by Vladimir N. Vapnik [29][30].

Let $S = \{(x_1, y_1), \ldots, (x_m, y_m)\}$ be a set of m training vectors. Each vector x_i is drawn from a domain $\mathscr{X} \subseteq \mathbb{R}^n$ and each label y_i is an integer from the set $\{-1, 1\}$. The idea is to find an hyperplane that separates the points x_i with label $y_i = -1$ from points x_i with label $y_i = 1$. This hyperplane represents the largest separation, or margin, between the two classes. So we choose the hyperplane so that the distance from it to the nearest data point on each side is maximized, as shown in Figure 1.9.

Linearly separable case.

Definition 1.1. The set S is said *linearly separable* if there exist $\omega \in \mathbb{R}^n$ and $b \in \mathbb{R}$ such that

$$y_i(\omega^T x_i + b) \geq 1 \qquad i = 1, \ldots, m \tag{1.9}$$

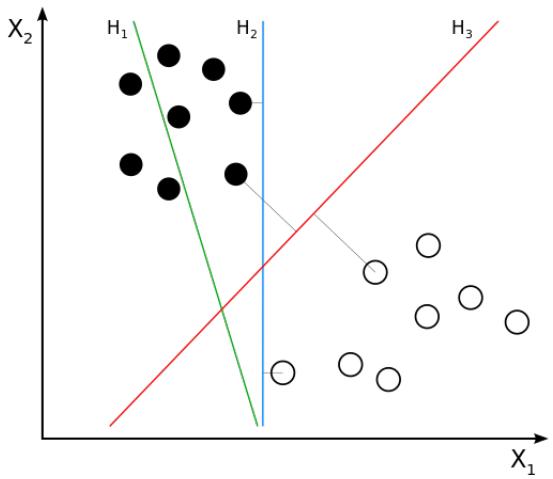


Figure 1.9: *Hyperplane H1 does not separate the classes. Hyperplane H2 does, but only with a small margin. Hyperplane H3 separates them with the maximum margin. Source: Wikipedia*

The pair (ω, b) defines a hyperplane of equation

$$\omega^T x + b = 0$$

named *separating hyperplane*.

In [31][32] is shown that to find the optimal hyperplane (i.e. the hyperplane with maximum separating margin), or equivalently, to determine a linear SVM, the following problem must be solve:

$$\begin{aligned} &\min_{\omega,b} \frac{1}{2}\|\omega\|^2 \\ &\text{s.t. } y_i(\omega^T x_i + b) - 1 \geq 0 \qquad i = 1, \ldots, m \end{aligned} \tag{1.10}$$

In order to solve problem (1.10) the dual formulation is introduced:

$$\begin{aligned} &\min_{\lambda} \sum_{i=1}^{m} \lambda_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} y_i y_j \lambda_i \lambda_j x_i x_j \\ &\text{s.t. } \sum_{i=1}^{m} \lambda_i y_i = 0 \\ &\qquad \lambda_i \geq 0 \qquad i = 1, \ldots, m \end{aligned} \tag{1.11}$$

The optimal solution λ^* of dual problem (1.11) allows to find the optimal hyperplane, as

$$\omega^* = \sum_{i=1}^{m} \lambda_i^* y_i x_i.$$

To find the scalar b^* the following KKT conditions must be considered:

$$\lambda_i^*[y_i((\omega^*)^T x_i + b) - 1] = 0 \qquad i = 1, \dots, m$$

and choosing a variable $\lambda_i^* \neq 0$ it is possible to compute b^* from:

$$y_i((\omega^*)^T x_i + b) - 1 = 0.$$

The vectors x_i closest to the optimal hyperplane (and correspondent to the dual variables λ_i^* different from zero) are called *support vectors*.

Linearly non-separable case. If the set S is not linearly separable it is possible to introduce m non-negative slack variables ξ_i such that

$$y_i(\omega^T x_i + b) \geq 1 - \xi_i \qquad i = 1, \dots, m$$

It is possible to modify the objective function of (1.10) such to penalize the slack variables different from zero:

$$\begin{aligned} &\min_{\omega, b, \xi} \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{m} \xi_i && \\ &\text{s.t. } y_i(\omega^T x_i + b) - 1 + \xi \geq 0 && i = 1, \dots, m \\ &\quad \xi_i \geq 0 && i = 1, \dots, m \end{aligned} \tag{1.12}$$

where C is a positive constant used as a regularization parameter, since the generalized optimal hyperplane tends to maximize the separating margin for small C and minimize the number of misclassified points for large C. The dual of this new formulation is similar to the previous one:

$$\begin{aligned} &\min_{\lambda} \sum_{i=1}^{m} \lambda_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} y_i y_j \lambda_i \lambda_j x_i x_j && \\ &\text{s.t. } \sum_{i=1}^{m} \lambda_i y_i = 0 && \\ &\quad 0 \leq \lambda_i \leq C && i = 1, \dots, m \end{aligned} \tag{1.13}$$

The difference is that the dual variables are now limited by C. To find the optimal ω^* and b^* it is possible to proceed in a similar way as in the separable case.

Non-linearly SVM. The previous solution for linearly non-separable case does not guarantee usually good performances since a hyperplane can only represent dichotomies in the space. For this reason, when the case is non linearly separable, the following strategy is used:

1. The input data are mapped into a higher dimensional space (called *feature space*).
2. The optimal hyperplane is computed (with the slack variables) in the feature space.

Formally, given the set S , let us consider the function

$$\phi : \mathbb{R}^n \to F$$

where F is the feature space, with dimension larger than m. The problem now is to determine the optimal hyperplane in the feature space F by substituting any occurrence of x_i with $\phi(x_i)$ in the dual problem (1.13), that become

$$\begin{aligned} &\min_{\lambda} \sum_{i=1}^{m} \lambda_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} y_i y_j \lambda_i \lambda_j \langle \phi(x_i), \phi(x_j) \rangle \\ &\text{s.t.} \sum_{i=1}^{m} \lambda_i y_i = 0 \\ &\quad 0 \leq \lambda_i \leq C \qquad\qquad i = 1, \ldots, m \end{aligned} \tag{1.14}$$

It is possible to define a function $K(\cdot,\cdot)$, called *kernel* such that

$$K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle.$$

If the kernel function satisfies Mercer's theorem [33] then it can represent a scalar product in the feature space, and the objective function of dual problem (1.14) can be rewritten as

$$\min_{\lambda} \sum_{i=1}^{m} \lambda_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} y_i y_j \lambda_i \lambda_j K(x_i, x_j).$$

It is possible to use different kernel functions [34]:

- Linear Kernel: it is given by the product plus a constant c:

$$K(x, y) = x^T y + c$$

- Polynomial Kernel:

$$K(x, y) = (a x^T y + c)^p$$

- Gaussian Kernel (or radial kernel): it maps samples into a higher dimensional space, so, unlike the linear kernel, it can handle the case when the relation between class labels and vectors is non linear

$$K(x,y) = e^{-\gamma\|x-y\|^2}$$

- Hyperbolic Tangent kernel:

$$k(x,y) = tanh(ax^T y + c)$$

Support Vector Regression (SVR)

ε-SVR. In ε-SVR the goal is to find a function $f(x)$ that has at most a deviation of ε from the actually obtained targets y_i for all the training data and at the same time is as flat as possible. In other words we do not care about errors as long as they are less than ε but will not accept any deviation larger than this value.

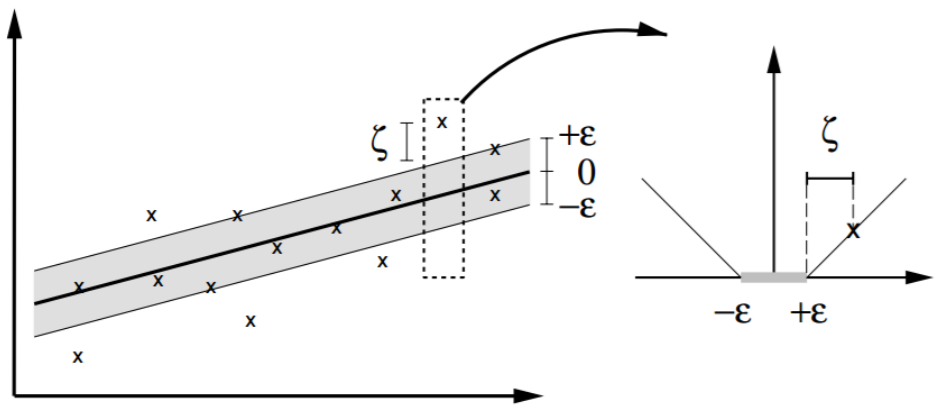


Figure 1.10: *ε-SVR.*

The related primal problem for this formulation is given by

$$\min_{\omega,b,\xi_i,\hat{\xi}_i} \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{N}(\xi_i + \hat{\xi}_i) \tag{1.15}$$

$$\begin{aligned} \text{s.t. } & \omega^T\phi(x_i) + b - y_i \leq \varepsilon + \xi_i && i = 1,\dots,m \\ & y_i - \omega^T\phi(x_i) - b \leq \varepsilon + \hat{\xi}_i && i = 1,\dots,m \\ & \xi_i, \hat{\xi}_i \geq 0 && i = 1,\dots,m \end{aligned}$$

The dual formulation of problem (1.15) is the following convex quadratic problem:

$$\begin{aligned}
&\min_{\lambda,\hat{\lambda}} \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}(\hat{\lambda}_i-\lambda_i)(\hat{\lambda}_j-\lambda_j)K(x_i,x_j)-\sum_{i=1}^{m}(\hat{\lambda}_i-\lambda_i)y_i+\varepsilon\sum_{i=1}^{m}(\hat{\lambda}_i+\lambda_i) \\
&\text{s.t. } \sum_{i=1}^{m}(\hat{\lambda}_i-\lambda_i)=0 \qquad i=1,\dots,m \\
&\quad 0\le\lambda_i,\hat{\lambda}_i\le C \qquad i=1,\dots,m
\end{aligned} \tag{1.16}$$

One difficulty with the ε-SVR is the selection of ε itself, which can become a hard work in absence of additional information about the noise contained in the training set, that is in general not available.

ν-SVR. To overcame with the problem of the selection of ε, in [35][36] has been introduced the ν-SVR formulation, which includes an additional term in the primal problem, a constant $\nu\ge 0$, which lets one control the number of support vectors and training errors. To be more precise, ν is an upper bound on the fraction of margin errors and a lower bound of the fraction of support vectors.

The primal formulation of the problem is

$$\begin{aligned}
&\min_{\omega,b,\xi_i,\hat{\xi}_i,\varepsilon} \frac{1}{2}\|\omega\|^2+C\big(\nu\varepsilon+\frac{1}{l}\sum_{i=1}^{m}(\xi_i+\hat{\xi}_i)\big) \\
&\text{s.t. } \omega^T\phi(x_i)+b-y_i\le\varepsilon+\xi_i \qquad i=1,\dots,m \\
&\quad y_i-\omega^T\phi(x_i)-b\le\varepsilon+\hat{\xi}_i \qquad i=1,\dots,m \\
&\quad \xi_i,\hat{\xi}_i\ge 0 \qquad i=1,\dots,m \\
&\quad \varepsilon\ge 0
\end{aligned} \tag{1.17}$$

while its dual has the following formulation

$$\begin{aligned}
&\min_{\lambda,\hat{\lambda}} \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}(\hat{\lambda}_i-\lambda_i)(\hat{\lambda}_j-\lambda_j)K(x_i,x_j)-\sum_{i=1}^{m}(\hat{\lambda}_i-\lambda_i)y_i+\varepsilon\sum_{i=1}^{m}(\hat{\lambda}_i+\lambda_i) \\
&\text{s.t. } \sum_{i=1}^{m}(\lambda_i-\hat{\lambda}_i)=0 \qquad i=1,\dots,m \\
&\quad 0\le\lambda_i,\hat{\lambda}_i\le\frac{C}{l} \qquad i=1,\dots,m \\
&\quad \sum_{i=1}^{m}(\lambda_i-\hat{\lambda}_i)\le C\nu \qquad i=1,\dots,m
\end{aligned} \tag{1.18}$$

The advantage of the ν-SVR is that ν can be chosen without prior knowledge of the noise present in the training data, and after the selection, the regression will chose ε to best fit this error, based on the requirements ν makes of the fraction of training-errors and support vectors.

References

[1] *About operations research*. INFORMS.org. Retrieved 7 January 2012.

[2] *What is OR*. HSOR.org. Retrieved 13 November 2011.

[3] F. S. Hillier and G. J. Lieberman. *Introduction to operations research*. McGraw-Hill Book Co, Singapore, 2001.

[4] W. E. Diewert. *Cost functions*. The New Palgrave Dictionary of Economics, 2nd Edition Contents, 2008.

[5] G. B. Dantzig. *Programming of interdependent activities: II mathematical model*. Econometrica, vol. 17, n. 3/4, pp. 200-211, 1949.

[6] G. B. Dantzig. *Linear programming under uncertainty*. Management Science, vol. 1, pp. 197–206, 1955.

[7] K. G. Murty. *Linear programming*. John Wiley & Sons, Inc. New York, 1983.

[8] M. Bartholomew-Biggs. *Nonlinear optimization with engineering applications*. Springer Optimization and Its Applications, vol. 19, 2008.

[9] M. Bartholomew-Biggs *Nonlinear optimization with financial applications*. Springer Science & Business Media, 2006.

[10] J. Borwein and Q. Zhu. *Techniques of variational analysis*. Springer. ISBN 978-1-4419-2026-3, 2005.

[11] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, Chapter 5, 2004.

[12] S. Rockafellar. *Convex analysis*. Princeton University Press, Chapter 28-30, 1970.

[13] J. Fortuny-Amat and B. McCarl. *A representation and economic interpretation of a two-level programming problem*. Journal of the Operational Research Society, vol. 32, n. 9, pp. 783-792, 1981.

[14] G. L. Nemhauser and L. A. Wolsey. *Integer and combinatorial optimization*. A Wiley-Interscience Publication, John Wiley & Sons, New York, Chapter II.4, 1988.

[15] S. K. Aggarwal, L. M. Saini and A. Kumar. *Electricity price forecasting in deregulated markets: a review and evaluation*. International Journal of Electrical Power and Energy System, vol. 31, n. 1, pp. 13-22, 2009.

[16] R. Weron. *Electricity price forecasting: a review of the state-of-the-art with a look into the future*. International Journal of Forecasting, vol. 30, n. 4, pp. 1030-1081, 2014.

[17] I. Moghram and S. Rahman. *Analysis and evaluation of five short-term load forecasting techniques*. IEEE Transaction on Power System, vol. 4, n. 4, pp. 1484-1491, 1989.

[18] S. M. Holland. *Principal component analysis (PCA)*. Department of Geology, University of Georgia, Athens, 2008.

[19] I. T. Joliffe. *Principal component analysis*. Springer, New York, 1986.

[20] H. F. Kaiser. *The application of electronic computers to factor analysis*. Educational and Psychological Measurement, vol. 20, pp. 141-151, 1960.

[21] U. E. Cali, B. Lange, J. Dobschinski, M. Kurt, C. Moehrlen and B. Ernst. *Artificial neural network based wind power forecasting using a multi-model approach*. 7th International Workshop on Large Scale Integration of Wind Power and on Transmission Networks for Offshore Wind Farms, 2008, Madrid.

[22] B. D. Ripley. *Pattern recognition and neural networks*. Cambridge University Press, 1996.

[23] L. Von Bremen. *Combination of deterministic and probabilistic meteorological models to enhance wind farm forecast*. Journal of Physics: Conference Series, vol. 75, 2007.

[24] J. Hertz, H. Krogh and R. G. Palmer. *Introduction to the theory of neural computation*. Lecture Notes, vol. 1, Santa Fe Institute, Addison-Wesley publishing company.

[25] S. Alessandrini, L. Delle Monache, S. Sperati and J. N. Nissen. *A novel application of an analog ensemble for short-term wind power forecasting*. Renewable Energy, vol. 76, pp. 768-781, 2015.

[26] L. Delle Monache, T. Nipen, Y. Liu, G. Roux and R. Stull. *Kalman filter and analog schemes to post-process numerical weather predictions*. Monthly Weather Review, vol. 139, pp. 3554–3570, 2011.

[27] S. Alessandrini, L. Delle Monache, S. Sperati and G. Cervone. *An analog ensemble for short-term probabilistic solar power forecast*. Applied Energy, vol. 157, pp. 95-110, November 2015.

[28] L. Delle Monache, A. Eckel, D. Rife and K. Searight. *Probabilistic weather predictions with an analog ensemble*. Monthly Weather Review, vol. 141, pp. 3498–3516, 2013.

[29] V. N. Vapnik. *The nature of statistical learning theory*. Springer, NY, 1995.

[30] V. N. Vapnik. *Statistical learning theory*. Wiley, NY, 1998.

[31] M. Sciandrone. *Support vector machine*. Lecture notes, University of Florence, 2005.

[32] J. Rosati. *Support Vector Machines for time series prediction*. Master Thesis, University of Camerino, 2012.

[33] N. I. Sapankevych and R. Sankar. *Time series prediction using support vector machine: a survey*. IEEE Computational Intelligence Magazine, vol. 4, pp. 24-38, 2009.

[34] B. Schölkopf and A. J. Smola. *Learning with kernels*. The MIT Press, 2000.

[35] B. Schölkopf, A. J. Smola and R. C. Williamson. *Shrinking the tube: a new support vector regression algorithm.* Advances in Neural Information Processing System, S. S. Kearns, S. A. Solla and D. A. Cohn (Eds), vol. 11, 1999.

[36] B. Schölkopf, A. J. Smola, R. C. Williamson and P. L. Bartlett. *New support vector algorithms*. Neural computation, vol. 12, pp. 1207-1245, 2000.

Chapter 2

Post-processing Techniques and Principal Component Analysis for Regional Wind Power and Solar Irradiance Forecasting

2.1 Introduction

In recent years, the development of wind and solar energy combined with their large-scale integration is creating a growing interest for predictions of the overall power production over large areas. In particular, Transmission System Operators (TSO) can benefit from such predictions to satisfy different purposes (e.g., overloads management or reserve estimation). The energy market operators need accurate predictions to reduce penalties, which are usually proportional to the unbalances, defined as the difference between predicted and observed power. Individual electricity producers are also interested in accurate predictions. Some of them have now penalties when they produce less than forecasted or have to manage storage in batteries to ensure a smoother export to the grid as part of their contract. Predicting renewable energy at large scale means aggregating the forecasts of the individual power plants located in a geographical area of interest. Such areas can be defined by a TSO, or they can correspond to energy market regions.

The purpose of this study is to test a forecasting approach for large areas with two different test cases. First, the approach is tested on all the wind farms located over Sicily (one of the Italian energy market regions). The hourly data of wind speed are available over 2 years (2011 and 2012). The installed capacity (IC) during this period is 1746 MW. The second test includes 98 Oklahoma Mesonet solar monitoring stations. In this case, the period considered is from January 2000 to December 2007, and the dataset is composed of the daily data of the aggregated solar radiation energy output.

Large-scale wind power and solar irradiance forecasts have been the subject of several studies [1][2]. The most straightforward method to predict the power generated over an entire area is to sum the individual forecasts of each power plant. To do that, the locations and characteristics of each plant must be known. In this chapter, an alternative method is proposed, which only requires historical time-series of power measurements produced by all the plants located within the area. In this study a Principal Component Analysis (PCA) technique is applied. PCA [3][4][5] is widely used in multivariate statistics and allows reducing the dataset size, retaining only the most relevant information. PCA is applied on the correlation matrix of the original forecasted variables over all the grid points inside the area, by computing its eigenvalues and eigenvectors. The eigenvalues allow establishing the optimal number of principal components, whereas the corresponding eigenvectors are used to transform the data, reducing the number of dimensions. A similar methodology was implemented to evaluate the performance of an analogous model on day-ahead forecasting of wind power production over large European regions in [6].

A Neural Network (NN) and an Analog Ensemble (AnEn) are applied on the PCA-transformed forecasts to obtain deterministic and probabilistic predictions of total wind power or total solar radiation energy. Probabilistic forecasting is useful to provide knowledge about uncertainty (e.g., a forecast of the PDF of power) in addition to a single-valued power prediction. For example, probabilistic predictions can be used by the day-ahead electricity markets for trading future energy production, obtaining higher results then those obtained by using deterministic forecasts alone, as shown in [7][8]. Indeed, by using probabilistic forecasting, it is possible to optimize the revenue for a producer in an economic model depending for example on the specific penalties for forecast errors valid in that market. The proposed method, AnEn [9][10], produces probabilistic power pre-

dictions from historical deterministic predictions and observations of the quantity to be predicted, as in [11][12][13][14][15]. It is possible to obtain a probabilistic power prediction also from Numerical Weather Prediction (NWP) ensembles [16][17][18]. The AnEn method was originally proposed in [19] for probabilistic meteorological forecasting. It searches for the past forecasts most similar to the current one and keeps the correspondent past observations to form an ensemble. From any probabilistic prediction, it is possible to obtain a single-valued deterministic prediction by computing the mean or the median of the n analog members.

In literature [20][21] it is possible to find different methods for deterministic power forecasting, that consists in a single value for each time in the future for the variable to be predicted. Reviews of the state-of-the-art are reported in [22][23]. A deterministic forecast can be generated with an artificial NN [24][25][26]. Given one or more inputs, e.g., the principal components, the NN produces a response variable (which in the present application consists in wind power or solar irradiance) through neural interconnections. The NN is initialized with numeric weights and its neurons can be trained to make them able to learn how to associate to each input the correspondent output. In this context, a single-layer feed-forward NN with a back-propagation algorithm is used.

The chapter is organized as follows: section 2.2 reports the observational and meteorological forecast data of Sicily and Oklahoma. The PCA technique is illustrated in section 2.3, within the post-processing methods AnEn and NN. Section 2.4 thoroughly describes the verification framework. Results of the application are given in section 2.5 and conclusions are illustrated is section 2.6.

2.2 Observational and meteorological data description

Hourly wind speed forecasts are generated over the area of Sicily using the Regional Atmospheric Modeling System (RAMS)[27] limited-area meteorological model. Initial and boundary conditions are obtained from the European Centre for Medium-Range Weather Forecasts (ECMWF) deterministic model [28].

The gridded wind field is obtained by interpolating data from the closest vertical model levels at 80 m a.g.l. which is the average hub height in Sicily. The whole compu-

tational domain is made of 900 grid points with 20 km horizontal grid increments. The RAMS simulation covers the period January 2011-December 2012, for which the hourly data of aggregated wind power produced in Sicily are made available by Terna, the Italian Transmission Operator (TSO) [29]. The average installed capacity (IC) during this period is equal to 1746 MW. We have applied a correction factor to account for an increasing IC during the overall period.

The first year of the dataset (2011) is used as a training period, whereas the second year (2012) is kept as an independent verification period. The training period is split into two sub-periods, of six months each; the first is used to choose the optimal number of principal components. We test this choice over the second part of the training dataset. Each forecast run is initialized at 12 UTC and covers the period from 0 to 72 hours ahead. We use a Geographic Information System (GIS) [30] to select from the whole domain only the 89 grid points within the territory of Sicily, as shown in Figure 2.1.

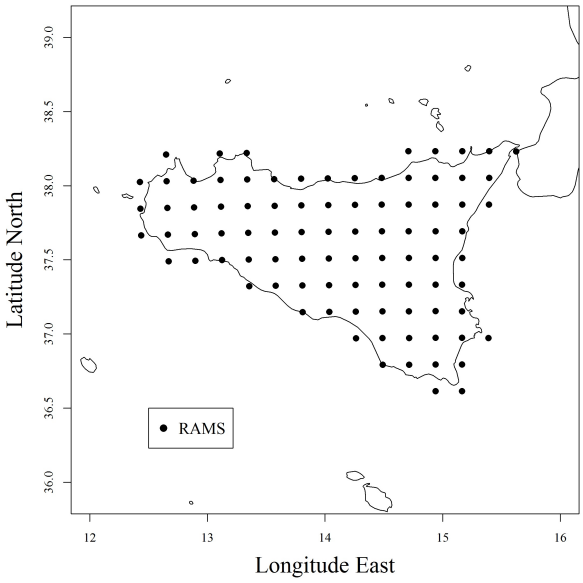


Figure 2.1: *Grid points by the RAMS model computational domain over Sicily (Italy), used for wind power predictions.*

The solar dataset was provided for a Kaggle competition [31]. Input NWP data for the contest are retrieved from the National Oceanic & Atmospheric Administration – Earth System Research Laboratory (NOAA/ESRL) Global Ensemble Forecast System (GEFS) Reforecast Version 2 [32]. Data include 11 ensemble members with three-hour

time resolution from 12 to 24 hours ahead. Each forecast run is initialized at 00 UTC. Locations of the Mesonet sites relative to the GEFS data ($1^\circ \times 1^\circ$ of spatial resolution) are shown in Figure 2.2.

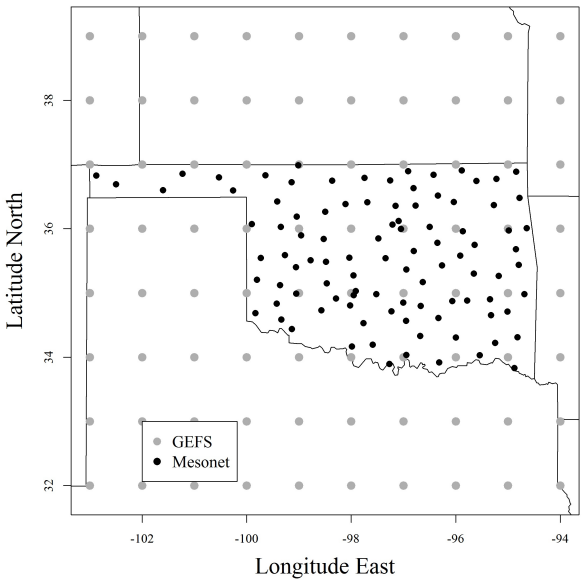


Figure 2.2: *NOAA/ESRL Global Ensemble Forecast System (GEFS) Reforecast Version 2 (big grey points) and Mesonet stations (small black points, see text for additional details) used for solar irradiance prediction.*

The total daily incoming solar radiation energy per square meter (referred as energy density, Jm^{-2} measured at 98 Oklahoma Mesonet sites is available since 1-January 1994. Solar radiation energy was measured by a pyranometer at each Mesonet site every five minutes and integrated from sunrise to 23:55 UTC of each day. For this work, only data of the period January 2000-December 2007 are considered. The maximum observed radiation energy density (MED) is equal to 22×10^8 Jm^{-2} in winter, 29×10^8 Jm^{-2} in spring and summer, and 21×10^8 Jm^{-2} in autumn. In this study we use only forecasts of the downward short-wave irradiative average flux at the surface cumulated on a daily basis (Jm^{-2}) to be consistend with the observations. These data are provided on 9×16 points grid covering Oklahoma, for a total of 144 points. Homogeneity across the different seasons is obtained by normalizing solar data on a daily basis. The normalization has been carried out by subtracting the mean and dividing by the standard deviation computed over the eight values corresponding to the same Julian day of the eight years in the dataset.

This is performed for both forecasted and measured data. The dataset is split into two parts of four years each: a training period (2000-2003) and a test period (2004-2007). The training period is split into two sub-periods of two years each with the same purpose as for the wind case.

2.3 Prediction methods

Principal component analysis (PCA)

PCA is a variable reduction procedure useful in case of datasets with a large number of correlated information. Principal components are linear combinations of observed variables, ordered according to a criterion of relevant information which is expressed by their variance: the first component extracts the maximum amount of information from the data and the other ones optimize the leftover information under the constraint of non-correlation with the other components. The primary purpose of this technique is to reduce the number of variables. This is achieved through a linear transformation of the m observed variables that are projected by the eigenvectors of the correlation matrix on a Cartesian system. The eigenvectors are sorted by descending order with respect to the amount of information; therefore, the variable with greater variance is projected on the first axis, the second on the second axis, and so on. To reduce the dataset dimension, the first $p \ll m$ components are taken.

To perform a PCA on a dataset, it is necessary to normalize the data by their mean and standard deviation to have more homogeneous distributions. Moreover, PCA constructs orthogonal, mutually uncorrelated, linear combinations that explain as much common variation as possible. PCA can be performed on both the covariance and the correlation matrix. Scaling the data matrix such that all variables have zero mean and unit variance (i.e., normalizing them) makes the two approaches identical. Let X be a $n \times m$ matrix of the normalized data, where n is the number of total forecast events used for the study, and m is the number of different grid points where the forecasts are available. In these applications $m = 89$ in the case of Sicily and $m = 144$ for Oklahoma.

The second step is to compute the eigenvalues and eigenvectors of the correlation

matrix. Let Λ denote the eigenvectors matrix. It is now possible to transform the original variables into new variables without information redundancy (uncorrelated) as follows:

$$X_{pca} = \Lambda \cdot X \tag{2.1}$$

After this operation, the original data are transformed in new uncorrelated data, without reducing the dimensionality. To reduce the data to a p–dimensional space, the original matrix X is multiplied by only the first p eigenvectors (called loadings). Each column of the resulting matrix, X_{pca} is a principal component, also known as score. In the literature three methods [3] have been introduced to choose the dimension p:

1. Cumulative percentage of total variation: only those components representing the $80\% - 90\%$ of the total variance are retained, i.e.,

$$\frac{\lambda_1 + \cdots \lambda_p}{\lambda_1 + \cdots \lambda_m} \approx 80\% - 90\% \tag{2.2}$$

   where λ_i is the i-th eigenvalues of the correlation matrix; the numerator represents the variance of the first p principal components, whereas the denominator represents the variance of all the main components.

2. The so-called Kaiser rule (proposed by Kaiser [33]) takes only those components having an eigenvalue greater than one.

3. Doing a "Scree Plot" where the number of principal components corresponds to the change of slope in the Scree Plot.

Since the first and the third methods depend on an arbitrary choice of the eigenvalues (i.e., there can be several choices of the number of eigenvalues), the second method is used in this work. In this study, the PCA is applied to wind speed gridded forecast data for the wind case and to the predicted gridded downward short-wave irradiative average flux at the surface for the solar case.

Starting from the 89 dimensions corresponding to the number of forecast grid points considered in Sicily, the PCA shows that only the first 6 principal components have eigenvalues greater than one; a sensitivity analysis on the number of components over the second part of the training dataset confirms this choice. For the solar case, the optimal number of principal components is 17.

Analog Ensemble (AnEn)

The AnEn method builds an ensemble forecast whose members are made by past observations appropriately selected across a historical dataset of deterministic meteorological forecasts. For these applications, the historical data-sets of forecasts are made by gridded data, the output of meteorological models. For the wind case, wind speed data from 89 points of RAMS simulation are used (see section 2.2 for more details). For the solar case, solar irradiance data from 144 points are provided by GEFS simulations.

Specifically, given a historical dataset of predictions and corresponding observations, for each lead-time of the current forecast, distances from every past forecast issued at the same lead-time are computed. All these distances are sorted from the lowest and the first n correspondent past forecasts are chosen. The concurrent n past measurements (wind power or solar radiation density in this work) are the n members that constitute the current analog ensemble forecast for the lead-time considered. It is important to underline that the AnEn, by selecting past power measurements which become the current forecast, naturally takes care of transforming meteorological predictions into power (or energy) predictions. The metric used to assess past forecasts similarity to the current forecasts is defined as follows [10][19][34]:

$$|F_t, A_{\bar{t}}| = \sum_{j=-\tau}^{j=\tau} \sqrt{\sum_{i=1}^{p} \lambda_i \left(F_{i,t+j} - A_{i,\bar{t}+j}\right)^2} \tag{2.3}$$

where F_t is the current forecast at time t at a certain location, $A_{\bar{t}}$ is an analog forecast at time $\bar{t}$ before F_t was issued and at the same location, λ_i is the i-th eigenvalue coming from the PCA, p is the dimension of the subspace in which the data are projected, τ is an integer equal to half width of the time window over which the metric is computed, and $A_{i,\bar{t}+j}$ and $F_{i,t+j}$ are the values of the analog and the forecast in the time window for a given variable. In this context, the forecast F and A are the arrays resulting from the PCA analysis. To assign different weights to each principal component, the eigenvalues λ_i are introduced in (2.3). Equation (2.3) differs from the metrics already used in [9] and [10]. In fact, in those works the forecasts F and A were not transformed by the PCA and represented arrays of different meteorological variables (predictors) extracted in one location. Also, λ_i were referred as the weights assigned to the different p meteorological predictors. In this work we just use one meteorological predictor (wind speed or solar irradiance) and

p represents the number of principal components selected. When the AnEn is applied without the PCA dimension reduction (see section 2.3), p is the number of grid points (89 for the wind case, 144 for the solar case) and λ_i is kept constant equal to 1.

The AnEn provides probability density functions (PDF) from which the likelihood of future events can be estimated. The PDF of the power or energy density forecast is estimated using the set of the n members from the analog ensemble forecast. On the other hand, computing the mean or the median of the n analog members allows obtaining a deterministic forecast as well.

Neural Network (NN)

A NN links the input variables and one or more output variables through an interconnected group of nodes called neurons. The back-propagation algorithm used in this work is based on the programming language R [35], and allows minimizing an error function depending on the different weights associated to the different interconnections. Each neuron assigns a weight to the input signals, and through an activation function, it produces an output term. The learning procedure of each neuron is performed through different iterative cycles, each making up a training phase for the network. At each iteration, the input signals in are multiplied by an appropriate value w representing the signal weight, usually within the range $(-1,1)$. This operation produces the resultant $net_i = \sum_j w_{ij} in_j$. The term $net_i - \theta_j$, where θ_j is an arbitrary value, forms the activation function argument and generates the output. The NN input is usually normalized to values between 0 and 1, to adapt the activation function to the weight values.

The strength of the learning phase is evaluated by the difference between the forecast output signal and the measured value; this difference represents the error. It is not possible to define a priori the optimal number of neurons, but in literature it is possible to find different rules [36][37]. The advantage of using an NN is that different types of data and variables, e.g., lead-time, wind speed and wind direction, can be combined. Also, being the training carried out using observed power data, it allows the NN to perform the conversion into power after getting meteorological variables as input. An independent iterative training is performed starting from different random weights: the NN that provides the minimum Root Mean Square Error (RMSE, see section 2.4), compared to

the observations over a subset of the training period, is selected and applied over the test period. In this work we have applied the NN using between 6 and 8 neurons depending on the different number of input data determined by each approach (see section 2.3). A sensitivity analysis (not shown) carried out on the training period has indicated the best NN configuration for each application.

Reference Methods

To better understand the PCA contribution to the overall forecast system's performance, we introduce a simpler dimension reduction technique as a baseline reference; we apply the AnEn and NN using the mean, the maximum, the minimum values and the standard deviation (MMMS) of the gridded forecast data as input. Also, we apply the AnEn and NN using the forecasts data from all the grid points as predictors. To summarize, we compare six approaches: NN after PCA (PCA+NN), AnEn after PCA (PCA+AnEn), NN after MMMS (MMMS+NN), AnEn after MMMS (MMMS+AnEn), and finally NN (NN) and AnEn (AnEn) without any dimension reduction.

We also want to understand the contribution of NN and AnEn as post-processing techniques to convert meteorological data into power data minimizing the BIAS. To do that, we compare the NN and AnEn after the PCA to a less sophisticated post processing method, i.e. a linear regression (PCA+LR). In Table 2.1 the computational time of PCA+NN, NN, PCA+AnEn and AnEn for both wind and solar cases are compared. The work has been implemented in R [35] and run on a single processor Intel® Core©i7-4510U, CPU 2.00 GHz.

Table 2.1: *Computational time for four different methods in both wind and solar case.*

	PCA+NN	NN	PCA+AnEn	AnEn
Wind	$3'\,41''$	$6h\ 39'\,44''$	$1h\ 50'\,6''$	$28h\ 4'\,35''$
Solar	$1'\,57''$	$18'\,7''$	$20'\,31''$	$3h\ 8'\,52''$

The improvement computational time reduction lead by PCA application is evident. When applied with the AnEn the reduction is of about 90% for both the wind and the solar case and when applied with the NN the reduction is of about 98% for the wind case and of about 88% for the solar case.

2.4 Forecast evaluation

Deterministic forecast evaluation

RMSE, Mean Absolute Error (MAE), bias and Pearson correlation indices are used to assess the performances of the different prediction systems. Bias is defined as the difference between the arithmetic mean of observed and forecasted values. It is obtained by the formula

$$bias = \frac{1}{N}\sum_{i=1}^{N}(f_i - o_i) \tag{2.4}$$

where f_i represents the forecasted data and o_i the measured one. The bias evaluates the systematic error and its optimal value is 0; a positive (negative) value indicates a tendency to overestimate (underestimate) the observations, respectively.

MAE is given by the formula

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|f_i - o_i| \tag{2.5}$$

i.e., it is the average absolute error between the forecasts and the observations. Lower MAE values indicate greater forecast accuracy.

RMSE is the squared root of the residual variance: it is given by the formula

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(f_i - o_i)^2} \tag{2.6}$$

i.e., it is the quadratic average of the errors and it indicates how much the forecasted data are far from the observed data. Similarly to MAE, a lower RMSE indicates greater accuracy.

The Pearson correlation index is a coefficient that assesses the linearity between the covariance of two random variables and the product of their standard deviation, based on the formula

$$cor = \frac{cov(o_i, f_i)}{\sqrt{sd(o_i)sd(f_i)}} \tag{2.7}$$

The range of this coefficient is $[-1,1]$. It evaluates how the forecast and the measurement temporal trend are similar. A correlation equal to 1 indicates a perfect forecast.
Bias, MAE and RMSE are normalized with respect to IC and MED.

Probabilistic forecast evaluation

The quality of probabilistic predictions can be evaluated by assessing several key attributes, some of which are described below.

Reliability and sharpness diagram

A reliability diagram allows visualizing the similarity between forecasts and observations in a probabilistic framework. The class intervals of forecast probabilities are plotted against the observed frequencies [38][39][40]. A probabilistic system is reliable if, for any given level of probability, the rate of occurrence of an event (i.e., its observed relative frequency) is equal to its forecasted probability. In other words, if an event is predicted with a certain probability, such event should occur in the same percentage of cases. A perfect probabilistic system would correspond to a reliability diagram with a $1:1$ diagonal line. Together with reliability it is also useful to show the sharpness of a probabilistic system, i.e., the ability to issue forecasts with both low and high probabilities [41]. By plotting the relative frequency (i.e., the number of cases) in each probability class interval, it is possible to assess how sharp the system is. Forecast systems with low sharpness exhibit a flat frequency line. For reliable forecasts sharpness leads to better resolution (i.e., predictive skill). Reliability diagrams, for each probabilistic system, are produced using a threshold equal to 50% of IC in the case of wind forecast and equal to 50% of seasonal MED in the case of solar forecasts, i.e., for each season the threshold is equal to 50% of maximum energy density of that season. Therefore four different thresholds are used for solar forecasts.

Relative Operating Characteristic Skill Score

The relative operating characteristic (ROC) skill score (ROCSS) index is based on the ROC curve [42], which is particularly useful in decision-making. The decision to

undertake actions about the occurrence of a particular event can depend on the forecasted probability of the same event exceeding a certain threshold. The ROC curve is produced by plotting the false alarms divided by the total number of non-occurred event against the hit rate (i.e., the correct forecast divided by the total occurrences of the event). A lower curvature corresponds to a higher number of false rates, showing a lower ability of the forecast system to discriminate. The ROC (as well as the ROCSS) thus depends upon resolution and not upon reliability. ROCSS translates the ROC score into a standard skill score estimating the resolution of probabilistic predictions, for which a value equal to 1 denotes a perfect forecast and a ROCSS lower than 0 indicates a system that performs worse than climatological forecasts.

Continuous Ranked Probability Score

The continuous ranked probability score (CRPS) compares a full probability distribution with the observations, when they are both represented as cumulative distribution functions (CDFs) [43]. It is given by the formula

$$CRPS = \frac{1}{N}\sum_{i=1}^{N}\int_{-\infty}^{+\infty}\left(F_i^f(x) - F_i^o(x)\right)^2 dx \tag{2.8}$$

where $F_i^f(x)$ is the CDF of the probabilistic forecast for the i-th values and $F_i^o(x)$is the CDF of the observations. The CRPS is a proper score and it is expressed in the same unit as the observed variable. A lower value of CRPS indicates better performance. The CRPS generalizes the mean absolute error [44] and therefore it provides a direct way of comparing various deterministic and probabilistic forecasts by using a single metric.

Rank histogram

Rank histogram is a diagnostic tool to evaluate the calibration and the consistency of an ensemble distribution, indicating how the ensemble spread represents the observation uncertainty [45]. The ensemble members are ranked to delineate ranges of the predicted variable; in a perfect ensemble the probability of occurrence of the observation within each range is equal. A calibrated system should have a flat distribution, in which the verification analysis ranks approximately with the same frequency in each interval. However,

in operational situations it is more frequently to observe different shapes of the distribution. When the distribution takes a U-form, it indicates an over-confident prediction: this means that most of the observations fall outside the extremes of the probability distribution, the ensemble members are not well calibrated and the spread is not sufficient to include all measurements. Conversely, an inverted U-shape indicates an under-confident estimate, in which the spread of the forecasts is excessively large: even in this case the members do not appear properly calibrated, since the observations fall mostly close to the center and not close enough to the extremes.

In the rank histograms, the horizontal line represents the frequency for a perfectly uniform distribution. Vertical confidence bars can be calculated with a quantile function for a binomial distribution, to show a range in which a deviation from the perfect distribution is still consistent with reliability. The bars delimit the 90% confident interval. The missing rate error (MRE), which is the fraction of observations lower (higher) than the lowest (highest) ranked prediction above or below the expected missing rate of $1/(n+1)$, is also shown. A larger positive (negative) MRE reveals a more underdispersive (overdispersive) ensemble.

Binned Spread/Skill Diagram

A binned-spread/skill diagram assesses the relationship between the ensemble spread and the RMSE of the ensemble mean over small intervals (i.e., bins) of spread, instead of considering its overall average; it assesses the ability of a probabilistic prediction to quantify its uncertainty [46][47][48]. Each bin has the same number of data points, which results in bins of different width. If the observations and the ensemble members are samples from the same distribution, it can be shown that the ensemble spread and the RMSE of the ensemble mean should be equal. This requirement becomes a necessary condition if the ensemble members and the ensemble mean errors do not have a Gaussian distribution. In this case, a good level of consistency is reached if the ensemble spread and the RMSE match at all values, resulting in a trend lying on the plot 1 : 1 diagonal. In this case the ensemble system is able to forecast its own error.

2.5 Results

Wind power results

The different methods are first compared in a deterministic framework. Figure 2.3 shows examples of the times series for PCA+NN, PCA+AnEn probabilistic forecasts, and the measured wind power. The PCA+AnEn quantile ranges are consistent with the wind power observations variability and, as expected, they increase for larger lead times.

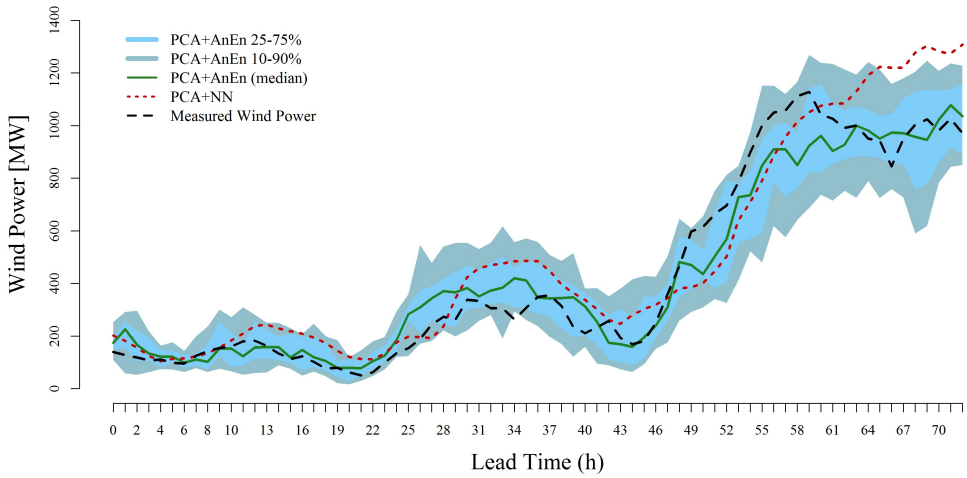


Figure 2.3: *PCA+NN, PCA+AnEn and measured wind power times series for 2 January 2012 for the wind case. The shadings correspond to the 25–75 (lighter) and 10–90 (darker) quantiles.*

Figure 2.4 shows RMSE (a), MAE (b), BIAS (c) (all normalized by IC) and correlation (d). Only lead-times from 36 to 59 hours ahead (i.e., the day-ahead) are shown (considering that lead time 0 corresponds to 12 UTC). All the forecast methods's performances deteriorate for lead times between 48-59 h.

This is a typical pattern caused by the presence of more variability in wind speed (hence less predictable conditions) due to stronger turbulent convection in the afternoon hours (i.e., lead times 48-59h). Using the PCA-reduction procedure instead of all the 89 grid points improve the results, in particular when the NN is applied. This is due to the high number of input that prevent the NN from working properly. In fact, with an increased number of degrees of freedom, it's more difficult to reach a convergence during the NN training. The AnEn can work with a high number of predictors (the results of PCA+AnEn, MMMS+AnEn and AnEn are comparable), but the advantage of using a

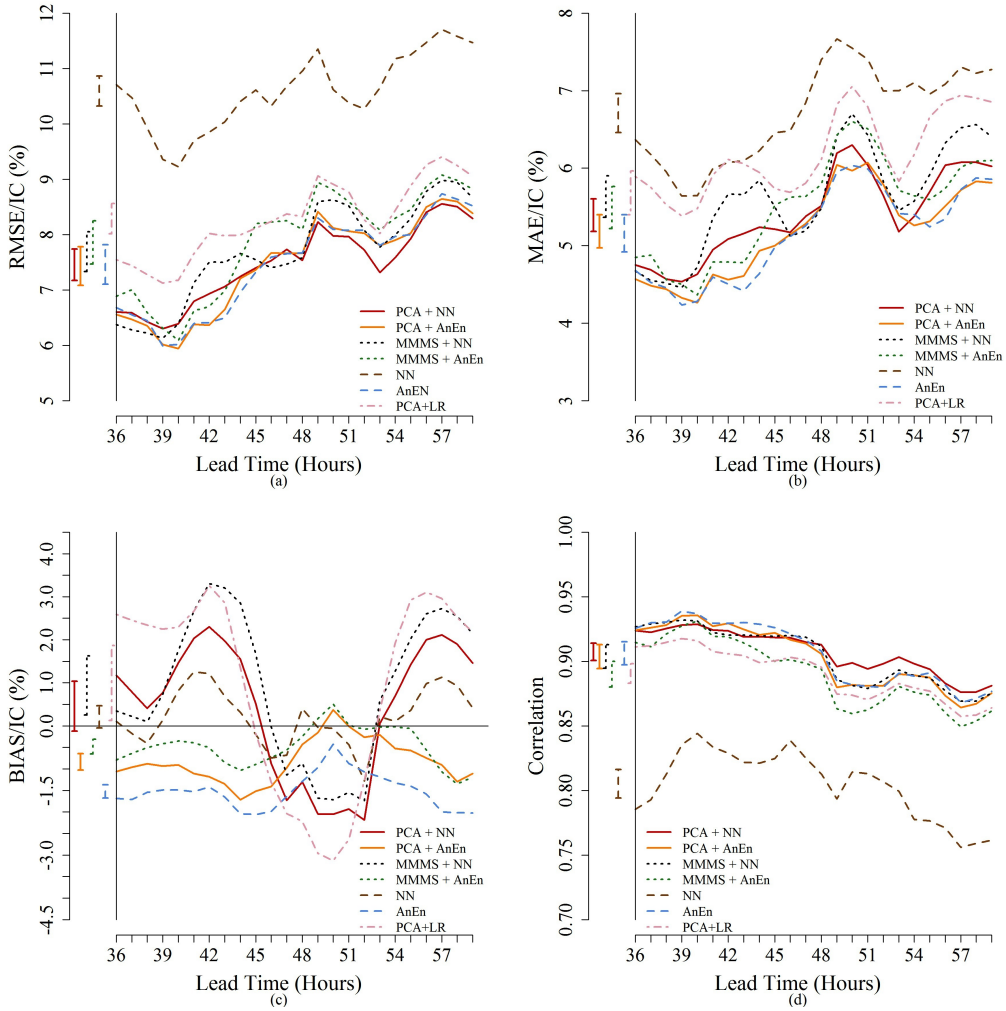


Figure 2.4: *RMSE (a), MAE (b), BIAS (c) and correlation (d) of PCA+NN, PCA+AnEn, MMMS+NN, MMMS+ AnEn,NN, AnEn and PCA+LR as a function of forecast lead time for 36-59h ahead for the wind case. The arrows to the left of each plot represent the bootstrap confidence intervals of each index obtained grouping data over the different lead times together.*

PCA relies in the significant computational time reduction (Table 1). When comparing the three different post-processing methods (NN, AnEn and LR), the simple LR performs worse than the NN and the AnEn. For the remaining part of the chapter we have compared only the NN and AnEn methods applied on the PCA-reduced data (i.e. PCA+NN, PCA+AnEn, MMMS+NN and MMMS+AnEn).. Using PCA instead of MMMS allows achieving better performances. PCA+NN shows generally slightly better performance in terms of RMSE, BIAS and correlation, whereas PCA+AnEn shows better performance in terms of MAE. NN performs slightly better than AnEn in terms of RMSE and correlation but, as shown by the confidence intervals in proximity of the left vertical axis (computed with all the lead times pulled together), the difference is not statistically significant. AnEn shows lower BIAS than NN.

Regarding the probabilistic evaluation, Figure 2.5 reports the reliability diagrams calculated using a threshold equal to 50% of IC corresponding to an observed frequency around 40% (see horizontal dashed line). The use of the PCA leads to a slightly more reliable forecast, in fact its forecast probabilities appear more similar to the observed frequencies in the range between 0.4 and 0.8. Regarding sharpness, both methods show similar results (i.e., the majority of cases occurs in the $0-20\%$ and $90-100\%$ range). The vertical bars in the diagrams are consistency intervals showing the range of empirical probabilities that could be observed even for perfect probabilistic forecast, due to sampling effects. They are derived from binomial distribution with the same probability as the corresponding class [49]. Bars are calculated using the 5% and 95% quantiles of this distribution. The horizontal dashed line is the climatological frequency (i.e., the mean of the observed frequencies).

Figure 2.6 (a) shows ROCSS values as a function of forecast lead time. The indices are computed using a threshold corresponding to 50% of IC. PCA+AnEn and MMMS+AnEn show a similar level of performance with values around 0.9 and always greater than 0.8 except for lead time 51-h. PCA+AnEn reaches higher values than MMMS+AnEn, in particular for lead time range 1-20. For all other lead times the two approaches are comparable. In Figure 2.6 (b) CRPS as a function of forecast lead-time is plotted. PCA+AnEn shows lower CRPS values for all of the lead times, indicating better performance. A daily cycle is also clear, with lower performances in early afternoon hours, when lower ROCSS

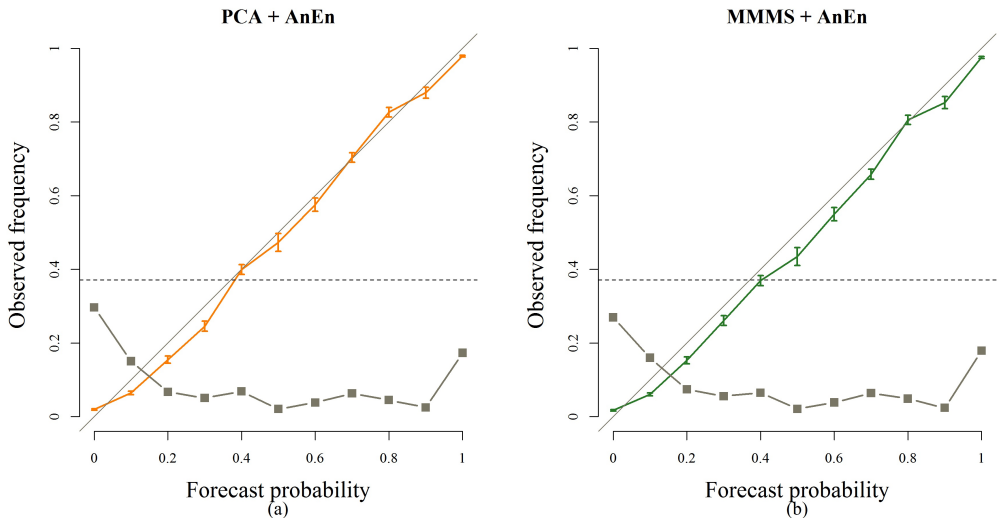


Figure 2.5: *Reliability (orange/green line) and sharpness (grey squares) diagrams for probabilistic forecasts of IC greater than 0.5, for PCA + AnEn (a) and MMMS + AnEn (b) for the wind case. The vertical bars indicate the potential range for perfectly reliable forecasts. The dashed horizontal line is the climatological frequency of the event.*

and higher CRPS values are observed.

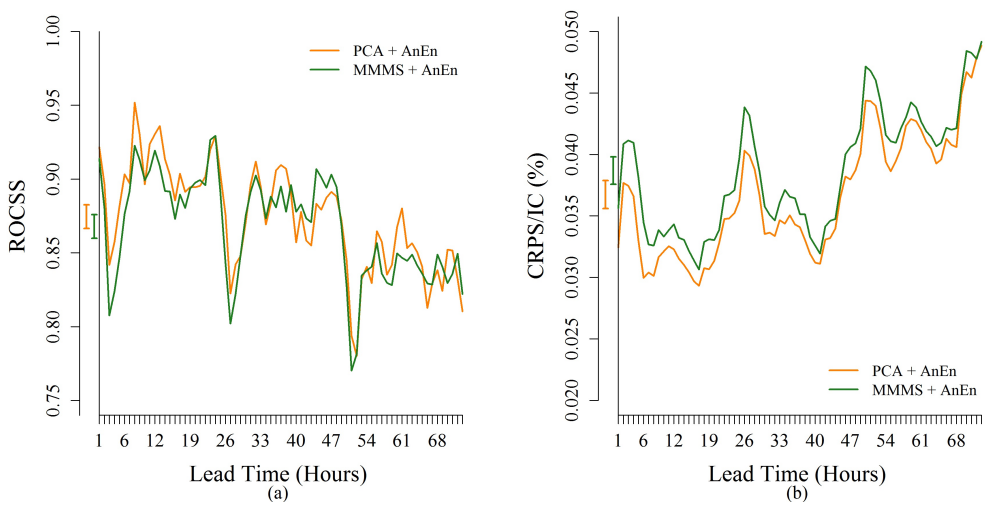


Figure 2.6: *ROCSS (a) as a function of forecast lead time (event forecasts greater than 0.5 IC) and CRPS (b) as a function of forecast lead time for the wind case. The orange line represents PCA + AnEn whereas the green one MMMS + AnEn. The arrows to the left of each plot represent the bootstrap confidence intervals of ROCSS and CRPS obtained by grouping data over the different lead-times together.*

Figure 2.7 shows the rank histogram produced for PCA+AnEn (a) and MMMS+AnEn (b). The horizontal line in the diagrams represents the frequency of a perfectly uniform

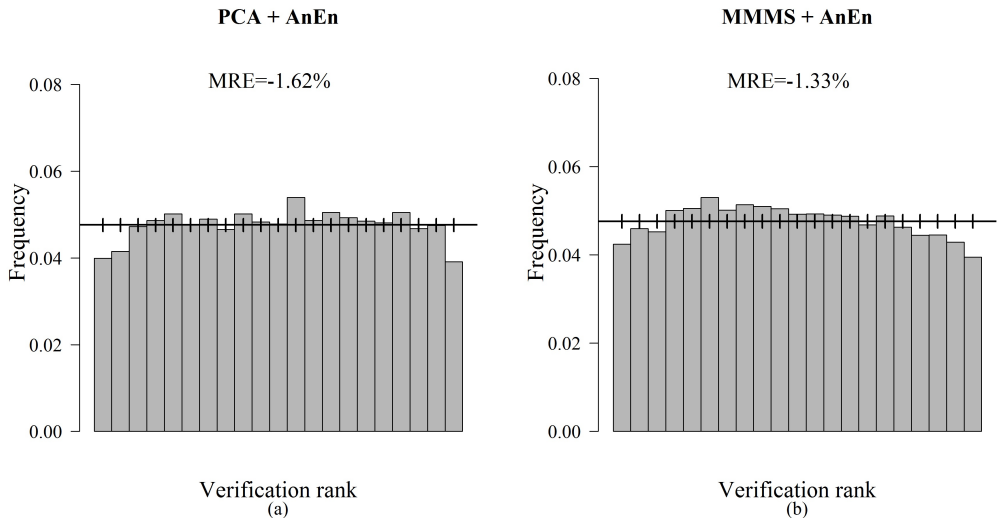


Figure 2.7: *Rank histogram, PCA + AnEn data (a), and MMMS + AnEn (b) for the wind case.*

distribution. The vertical bars delimit the 90% confidence interval. Both methods appear slightly overdispersive with very similar rank histograms, with small MRE values and a good statistical consistency.

We perform an additional assessment of statistical consistency by compiling binned - spread/skill diagrams as in Figure 2.8. The two different approaches show a good distribution of their values along the diagonal, even if they are both slightly overdispersive. The two approaches are both adequately able to represent forecast uncertainty, since their spread reflects the deterministic error variance.

Solar irradiance results

Figure 2.9 shows time series of forecasted solar irradiance for PCA+NN and PCA+AnEn and the observed values. The forecasted values look in good agreement with the observations.

As shown in Figure 2.10, for deterministic predictions, a reduction procedure (PCA or MMMS) leads to better results than the use of the entire dataset, in particular for the NN. The AnEn also performs worse when applied to all the 144 grid points exhibiting worse BIAS values. Concerning the post-processing methods, the NN and the AnEn perform better than the LR. As for the wind case, in the remaining part of the chapter we have tested the PCA+NN and the PCA+AnEn only. The PCA approach leads to better results

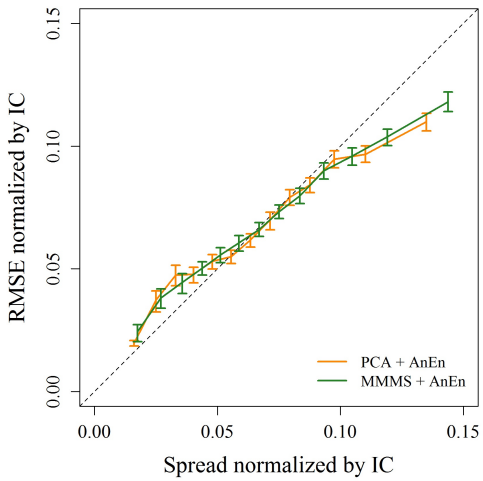


Figure 2.8: *Binned ensemble spread versus standard deviation of the ensemble mean error for the probabilistic forecasts: PCA + AnEn (orange line) and MMMS + AnEn (green line) for the wind case. The bars delimit the* 90% *bootstrap confidence interval.*

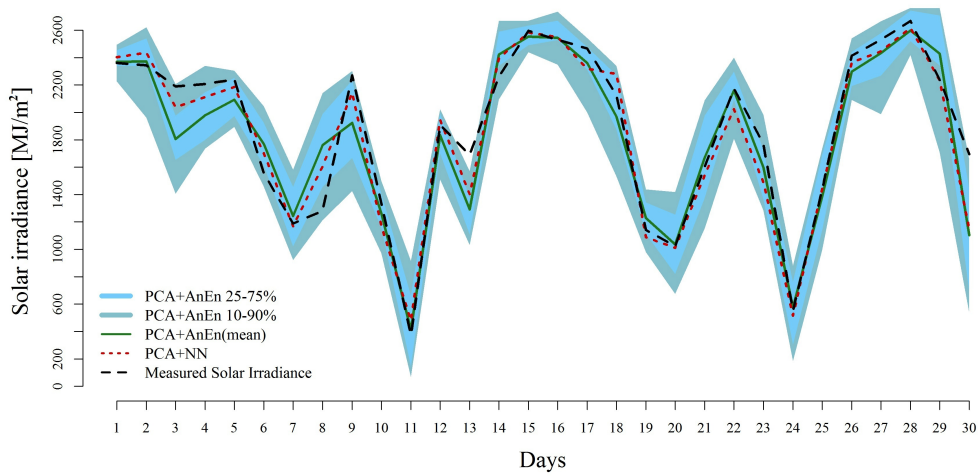


Figure 2.9: *PCA+NN, PCA+AnEn and measured solar irradiance times series for April 2004 for the solar case. The shadings correspond to the 25–75 (lighter) and 10–90 (darker) quantiles.*

than MMMS in terms of RMSE (a), MAE (b), BIAS (c) (normalized by the maximum energy density) and correlation (d). The trends are similar for all four cases, showing good results during winter and autumn seasons and lower performances during spring and summer. Indeed, prediction errors are proportional to the solar radiation energy that is lower during winter [34]. NN leads to slightly better results than AnEn, both coupled to PCA and MMMS. RMSE and MAE follow the same trend, showing higher performance

for NN, in particular for PCA+NN. BIAS shows a tendency to underestimate the forecasts in spring summer and autumn. Looking at the bootstrap intervals (in proximity of the left vertical axis computed with all the lead times pulled together), the improvement in correlation on PCA+NN and PCA+AnEn is closer to be statistically significant than for other metrics.

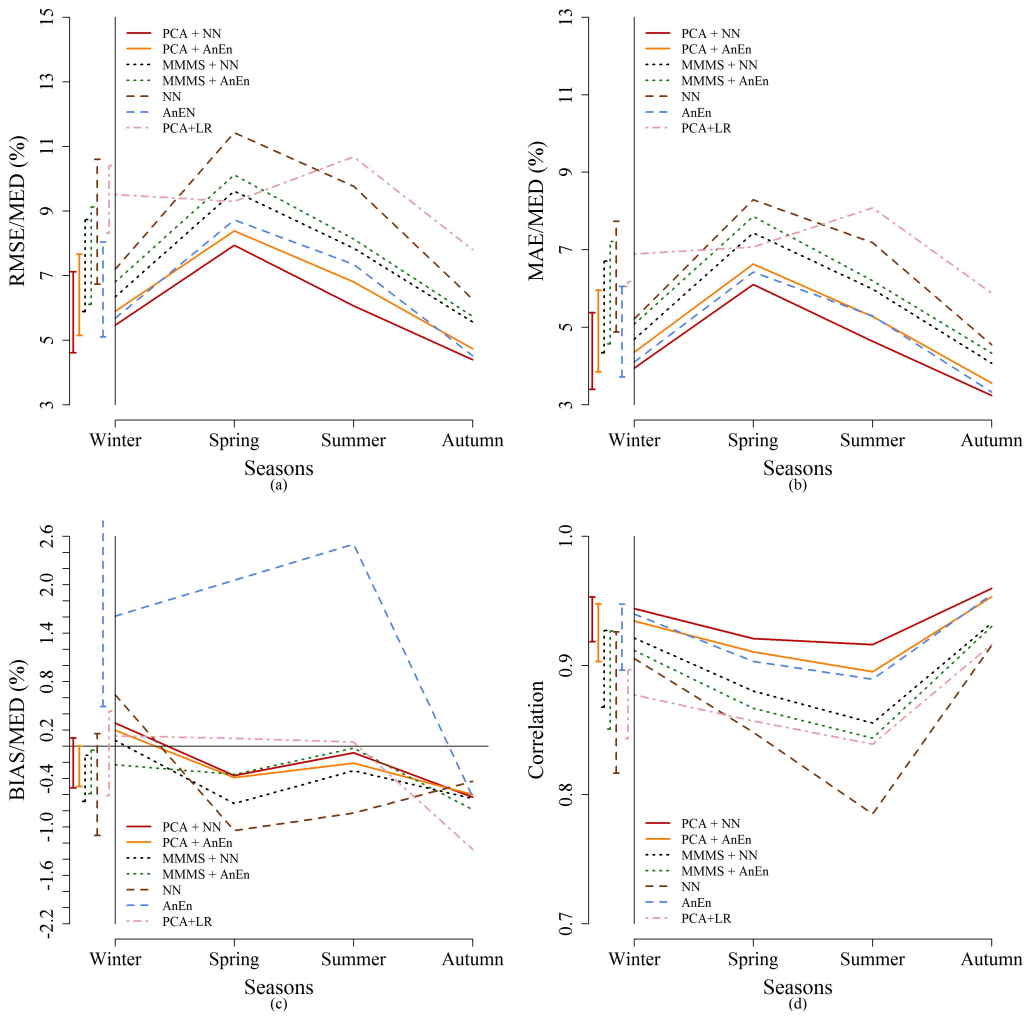


Figure 2.10: *RMSE (a), MAE (b), BIAS (c) and correlation (d) of PCA+NN, PCA+AnEn, MMMS+NN, MMMS+ AnEn,NN, AnEn, PCA+LR as a function of the seasons for the solar case. The arrows to the left of each plot represent the total value of each index computed on the entire dataset with the bootstrap confidence interval.*

In Figure 2.11 PCA+AnEn (a) and MMMS+AnEn (b) are compared by plotting reliability diagrams, where the threshold is set to a value equal to 50% of MED. The two models show the same level of reliability and sharpness. The horizontal dashed line is the

climatological frequency around 55%.

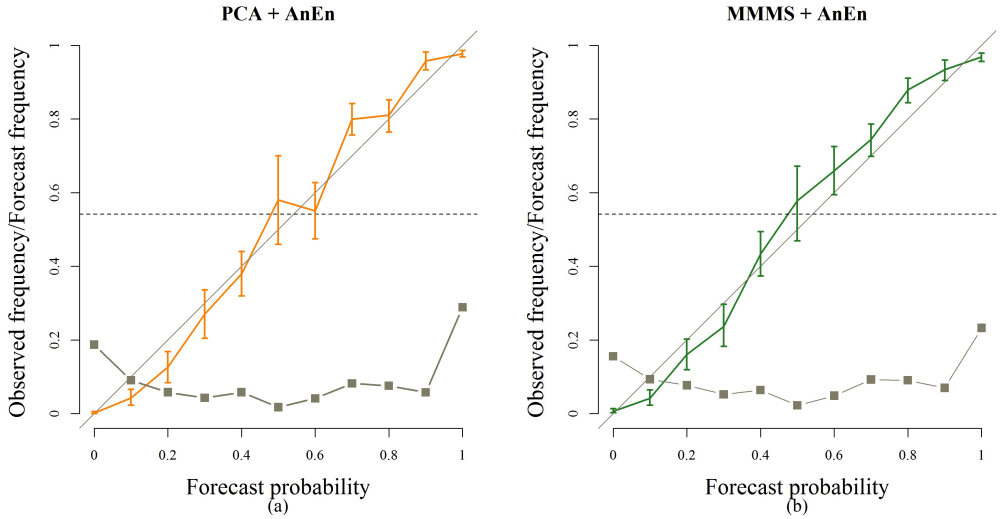


Figure 2.11: *Reliability (orange/green line) and sharpness (grey squares) diagrams (event forecast greater than 0.5 MED), for PCA + AneEn (a) and MMMS + AnEn (b) for the solar case. The vertical bars indicate the potential range for perfectly reliable forecasts. The dashed horizontal line is the climatological frequency of the event.*

Figure 2.12 shows ROCSS (a) and CRPS (b) diagrams, where the corresponding values are reported as a function of the season. In the ROCSS diagram the indices are computed for a threshold corresponding to 50% of MED. PCA+AnEn shows a level of performance with values always greater than 0.9, whereas MMMS+AnEn shows lower values, with a minimum around 0.7 in summer. The bootstrap intervals indicate that the differences are statistically significant only during summer. This means that PCA+AnEn performs statistically significantly better than MMMS+AnEn only during summer. PCA+AnEn model shows lower CRPS values than MMMS+AnEn during all the seasons. The rank histograms and binned spread-skill diagrams (not shown) for both PCA+AnEn and MMMS+AnEn show an optimal and similar level of statistical consistency.

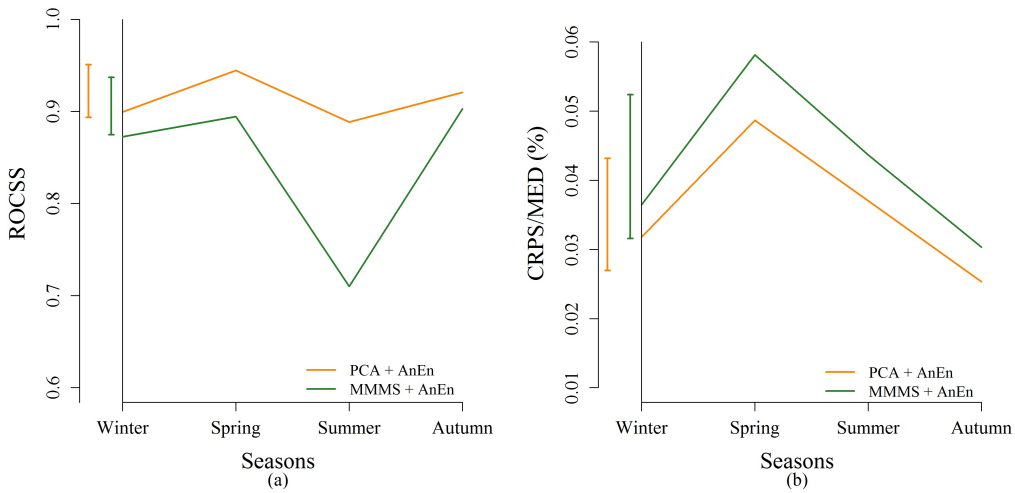


Figure 2.12: *ROCSS (a) as a function of seasons for the probabilistic forecasts of forecasted energy density greater than 0.5 MP and CRPS (b) a function of seasons for the solar case. The orange line represents PCA + AnEn whereas the green one MMMS + AnEn. The arrows to the left of each plot represent the total value of each index computed on the entire dataset with the bootstrap confidence interval.*

2.6 Conclusions

We test different approaches to predict the hourly/daily wind power or solar irradiance produced by several plants spread over a vast geographic region on two separate datasets: Sicily (South of Italy) for wind power and Oklahoma for solar irradiance. These approaches are based on using gridded forecast data from Numerical Weather Prediction (NWP) models over the geographic region of interest after a Principal Component Analysis (PCA). We use the PCA to reduce the size of the forecast gridded data. After the PCA, starting from the dimensionally reduced datasets, we apply the Analog Ensemble (AnEn) technique and a Neural Network (NN) algorithm to generate deterministic and probabilistic predictions of the regional power output.

In both the solar (Oklahoma) and wind power (Sicily) applications, the PCA-based reduction procedure leads to better results than those obtained by applying the NN and AnEn directly over the entire gridded prediction data. The improvements rely both on the forecast accuracy (lower Root Mean Squared Error, RMSE) and on consistent computational time reduction (between 88% and 98% depending on the application). Also, AnEn and NN both outperform a simpler bias reduction method as a linear regression.

We also compared the PCA with a less sophisticated dimension reduction procedure, as taking the minimum the maximum and the mean values of the gridded forecast data (MMMS) For Sicily, PCA leads to better results than MMMS, particularly when coupled with NN. RMSE, MAE, BIAS, and correlation show that PCA performs better than MMMS, which leads to a lower performance possibly caused by a loss of information. Indeed, the selected principal components contain a large portion of the available information (described by the variance) of the original data. Thus, they are a valid proxy of the entire dataset.

For probabilistic wind power predictions over Sicily, the MMMS and PCA are almost similar,even though MMMS shows a tendency to overestimate the forecasts and to be less reliable than PCA. The benefit of using a PCA-based approach is more evident for the Oklahoma dataset. When NN s applied after PCA, it provides the most accurate deterministic forecast, particularly during summer. PCA+AnEn is more reliable than MMMS+AnEn when used to generate probabilistic solar power predictions. Also, lower values of CRPS and higher values of relative operating characteristic skill score (ROCSS) indicate a higher quality of the PCA-based probabilistic forecasts.

Notes

This chapter has been published as article on Solar Energy 134, 2016, pages 327–338. Received 27 January 2016, received in revised form 7 April 2016, Accepted 28 April 2016. The co-authors are Stefano Alessandrini (NCAR), Simone Sperati (RSE), Luca Delle Monache (NCAR), Davide Airoldi (RSE), Maria Teresa Vespucci (University of Bergamo). NCAR is National Centre for Atmospheric Research, Boulder, CO, USA.

References

[1] R. Girard, K. Laquaine and G. Karioniotakis. *Assessment of wind power predictability as a decision factor in the investment phase of wind farms*. Applied Energy, vol. 101, pp. 609-617, 2013.

[2] P. J. Trombe, P. Pinson and H. Madsen. *A general probabilistic forecasting framework for offshore wind power fluctuations*. Energies, vol. 5, pp. 621-657, 2012.

[3] I. T. Joliffe. *Principal component analysis*. Springer, New York, 1986.

[4] S. Wold, K. Esbensen and P. Geladi. *Principal component analysis*. Chemometrics and Intelligent Laboratory Systems, vol. 2, pp. 37-52, 1987.

[5] H. Abdi and L. J. Williams. *Principal component analysis*. vol. 2, pp. 433-459, 2010.

[6] M. L. Martin, F. Valero, A. Pascual, J. Sanz and L. Frias. *Analysis of wind power production by means of an analog model*. Atmospheric Research, vol. 143, pp. 238–249, 2014.

[7] M. S. Roulston, D. T. Kaplan, J. Hardenberg and L. A. Smith. *Using medium-range weather forecasts to improve the value of wind power production*. Renewable Energy, vol. 28, pp. 585-602, 2002.

[8] S. Alessandrini, F. Davò, S. Sperati, M. Benini and L. Delle Monache. *Comparison of the economic impact of different wind power forecast system for producers*. Advances in Science Research, vol. 11, pp. 49-53, 2014.

[9] S. Alessandrini, L. Delle Monache, S. Sperati and J. N. Nissen. *A novel application of an analog ensemble for short-term wind power forecasting*. Renewable Energy, vol. 76, pp. 768-781, 2015.

[10] L. Delle Monache, T. Nipen, Y. Liu, G. Roux and R. Stull. *Kalman filter and analog schemes to post-process numerical weather predictions*. Monthly Weather Review, vol. 139, pp. 3554–3570, 2011.

[11] J. B. Bremnes. *A comparison of a few statistical models for making quantile wind power forecasts*. Wind Energy, vol. 9(1-2), pp. 3-11, 2006.

[12] J. K. Møller, H. A. Nielsen and H. Madsen. *Time-adaptive quantile regression*. Computational Statistics & Data Analysis, vol. 52, pp. 1292–1303, 2006.

[13] H. A. Nielsen, H. Madsen and T. S. Nielsen. *Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts*. Wind Energy, vol. 9(1-2), pp. 95-108, 2006.

[14] J. Wang, A. Botterud, R. Bessa, H. Keko, L. Carvalho, D. Issicaba, J. Sumali and V. Miranda. *Wind power forecasting uncertainty an unit commitment*. Applied Energy, vol. 88, pp. 1014-1023, 2011.

[15] P. Pinson. *Estimation of the uncertainty in wind power forecasting*. Ph.D. dissertation, Ecole des Mines de Paris, 2006.

[16] S. Alessandrini, S. Sperati and P. Pinson. *A comparison between the ECMWF and COSMO Ensemble Prediction System applied to short-term wind power forecasting on real data*. Applied Energy, vol. 107, pp. 271-280, 2013.

[17] S. Alessandrini, P. Pinson, R. Hagedorn, G. Decimi and S. Sperati. *An application of ensemble/multi model approach for wind power production forecasting*. Advances in Science Research, vol. 6, pp. 35-37, 2011.

[18] M. Lange and U. Focken. *Physical approach to short-term wind power forecast*. Springer, 2005.

[19] L. Delle Monache, A. Eckel, D. Rife and K. Searight. *Probabilistic weather predictions with an analog ensemble*. Monthly Weather Review, vol. 141, pp. 3498–3516, 2013.

[20] F. Cassola and M. Burlando *Wind speed and wind energy forecast through Kalman filtering of numerical weather prediction model output*. Applied Energy, vol. 99, pp. 154–166, 2012.

[21] P. Pinson, H. A. Nielsen, H. Madsen and T. S. Nielsen. *Local linear regression with adaptive orthogonal fitting for the wind power application*. Statistics and Computing, vol. 18, pp. 59-71, 2008.

[22] G. Giebel, R. Brownsword, G. Kariniotakis, M. Denhard and C. Draxl. *The state-of-the-art in short-term prediction of wind power: a literature overview*. ANEMOS plus, 2011.

[23] S. Sperati, S. Alessandrini, P. Pinson and G. Karioniotakis. *The "weather intelligence for renewable energies" benchmarking exercises on short-term forecasting in wind and solar power generation*. Energies, vol. 8(9), pp. 9594-9619, 2015.

[24] U. E. Cali, B. Lange, J. Dobschinski, M. Kurt, C. Moehrlen and B. Ernst. *Artificial neural network based wind power forecasting using a multi-model approach*. 7th International Workshop on Large Scale Integration of Wind Power and on Transmission Networks for Offshore Wind Farms, 2008, Madrid.

[25] B. D. Ripley. *Pattern recognition and neural networks*. Cambridge University Press, 1996.

[26] L. Von Bremen. *Combination of deterministic and probabilistic meteorological models to enhance wind farm forecast*. Journal of Physics: Conference Series, vol. 75, 2007.

[27] R. A. Pielke, W. R. Cotton, R. L. Walko et al. *A comprehensive meteorological modeling system – RAMS*. Meteorology and Atmospheric Physics, vol. 49, pp. 69-91, 1992.

[28] ECMWF. *European Centre for Medium-Range Weather Forecasts*. 2015, http://www.ecmwf.int/.

[29] Terna. http://www.terna.it/default.aspx?tabid=1778.

[30] QGIS. http://www2.qgis.org/it/site/about/index.html.

[31] Kaggle. *Kaggle competion*, 2014. https://www.kaggle.com/c/ams-2014-solar-energy-prediction-contest/data.

[32] T. Hamill, G. Bates, J. Whitaker, D. Murray, M. Fiorino and T. Galarneau. *A description of the 2nd generation NOAA global ensemble reforecasts data set*. NOAA Earth System Research Lab, Physical Sciences Division Boulder, 2015.

[33] H. F. Kaiser. *The application of electronic computers to factor analysis*. Educational and Psychological Measurement, vol. 20, pp. 141-151, 1960.

[34] S. Alessandrini, L. Delle Monache, S. Sperati and G. Cervone. *An analog ensemble for short-term probabilistic solar power forecast*. Applied Energy, vol. 157, pp. 95-110, Nov. 2015.

[35] R Foundation for Statistical Computing. *R, A language and environment for statistical computing*. http://www.R-project.org.

[36] Z. Boger and H. Guterman. *Knowledge extraction from artificial neural network models*. IEEE Systems, Man, and Cybernetics Conference, 1997.

[37] A. Blum. *Neural networks in C++*. Wiley, 1992.

[38] I. T. Jolliffe and D. B. Stephenson. *Forecast verification: a practitioner's guide in atmospheric science*. Wiley, 2003.

[39] D. S. Wilks. *Statistical methods in the atmospheric sciences*. International Geophysics, 2006.

[40] P. Pinson, P. McSharry and H. Madsen. *Reliability diagrams for nonparametric density forecasts of continuous variables: accounting for serial correlation*. Quarterly Journal of the Royal Meteorological Society, vol. 136, pp. 77-90, January 2010.

[41] T. Gneiting, A. E. Raftery, F. Balabdaoui and A. Westveld. *Verifying probabilistic forecasts: calibration and sharpness*. Proc Workshop on Ensemble 670 Weather Forecasting in the Short to Medium Range, 2003.

[42] I. B. Mason. *A model for the assessment of weather forecasts*. Australian Meteorological Magazine, vol. 30, pp. 291–303, 1982.

[43] E. P. Grimit, T. Gneiting, V. Berrocal and N. A. Johnson. *The continuous ranked probability score for circular variables and its application to mesoscale forecast ensemble verification*. Technical report no. 493, Department of statistics, university of Washington, 2006.

[44] M. Carney and P. Cunningham. *Evaluating density forecasting models*. Technical report, Trinity College Dublin, Department of computer Sciences, 2012.

[45] T. Hamill. *Interpretation of rank histograms for verifying ensemble forecasts*. Monthly Weather Review, vol. 129, pp. 550-560, 2001.

[46] H. M. Van Den Dool. *A new look at weather forecasting through analogues*. Monthly Weather Review, vol. 117, pp. 2230-2247, 1989.

[47] X. Wang and C. H. Bishop. *A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes*. Journal of Atmospheric Science, vol. 60(9), pp. 1140–58, 2003.

[48] T. M. Hopson. *Assessing the ensemble spread-error relationship*. Monthly Weather Review, vol. 142(3), pp. 1125–42, 2014.

[49] J. Bröcker and L. A. Smith. *Increasing the reliability of reliability diagrams*. Weather and Forecasting, vol. 22, pp 651-661, 2007.

Chapter 3

Forecasting Italian electricity market prices using a Neural Network and a Support Vector Regression

3.1 Introduction

Electricity price forecasting has always been an important tool in a competitive market. Participants in deregulated electricity market can use price forecasting to develop their bidding strategies to maximize the profit obtained by trading energy. In bilateral contracts the agreed price of buyer and seller is based on market clearing price predictions. On the other hand, the market behaviour can be analysed by the market operators using accurate price forecasts.

Price forecasting has been the subject of several studies (see e.g. [1] and [2] for a review), and a lot of different methods have been proposed for price and load forecasting, such as linear regressions [3], stochastic processes [4], ARMA models [5], and also weighted nearest neighbours techniques [6], quantile regressions [7][8] or hybrid correction method [9]. To obtain excellent results in price forecasting it is not sufficient to have good methods for deriving such forecasts: it is very important to have accurate information. Indeed, electricity demand (and therefore electricity prices) strictly depends on weather (e.g., temperature, wind speed, precipitation) and on the peak differences during a day or a week. Furthermore, since electricity is non-easily storable and power systems

require a constant balance between production and consumption, it is important to have information more precisely as possible. The problem is that this information is rarely available to individual market participants.

The purpose of this study is to test the role of different predictors, in particular those easy achievable (e.g., on websites of Power Exchange Operators - PEO, or Transmission System Operators - TSO). Furthermore, we want to test the role that each predictor has in different price forecasting approaches. In particular the study is conducted over 2 years (2014 - 2015) for the Italian day ahead electricity market prices, where supply orders, pumping demand orders and export demand orders are driven by the zonal prices, while all the other demand orders are driven by an Uniform Purchase Price (Prezzo Unico Nazionale or PUN) [10]. We test our predictors and methods both on PUN and on some zonal prices, in particular those of zones called Nord, Centro-Nord, Centro-Sud and Sud, to try to understand which factors are crucial in price forecasting and which are not so decisive. The decision to forecast both PUN and zonal prices is due to the fact that PUN is the price at which most of the Italian demand orders are subject and it is unique for all the Italian electricity market; zonal prices, instead, differ from a zone to another when there is a congestion between these two zones (the zones Nord and Centro-Nord have had a different price in the 20% of the hours in 2015, the percentage rises to 34% comparing the zones Nord and Sud). Therefore it is important to classify the most important information between zones. We forecast the prices of the day-ahead electricity market, that means that, if we are in day D, we forecast hours of day $D+1$.

Since we want to work with a big variety of data, in this chapter we proposed two different approaches able to work with a high number of different predictors: a Neural Network (NN) and a Support Vector Regression (SVR).

The NN has been adopted thanks to its ability to learn how to associate to each input the correspondent output, and in literature is possible to find different examples (see, e.g., [11] [12][13]). Given different predictors, the NN produces the forecasted PUN (or forecasted zonal prices) through neural interconnections. To achieve this purpose the neurons of the NN must be trained.

SVR has been used for regression and time series prediction, as in [14][15], and it is based on Support Vector Machine (SVM), a machine learning method that has been

recently introduced in the framework of statistical learning theory (developed by Vapnik [16][17]) for data classification. SVM are usually used for time series forecasts in particular in presence of non-linearity. The strength of SVM is that it maps the input data into a high dimensional space and then uses simple linear functions to create linear constraints in the new space. For these reasons we thought that SVR can be a good method for our purposes.

The novelty proposed in this chapter is that we tried to apply NN and SVR to a big number of different and easily achievable predictors.

The chapter is organized as follows: Section 3.2 reports a description of all the available data that have been used in the model. The predictions methods NN and SVR are illustrated in Section 3.3. Results of the application (both for PUN and zonal prices) are given in Section 3.4 and conclusions are illustrated in Section 3.5.

3.2 Data Description

Hourly prices data cover the period January 2014 – December 2015. We chose to start the dataset from January 2014 and not to use the years before since the Italian electricity price has changed during the last years, in particular prices of current days are lower than prices before 2014, as is shown for PUN in Figure 3.1 (the same holds for zonal prices).

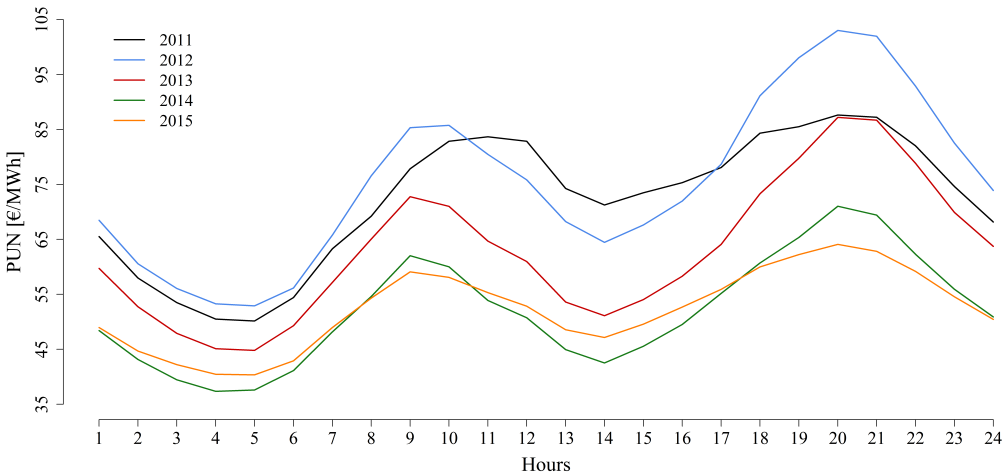


Figure 3.1: *Hourly mean of PUN as function of the hours for five different years.*

PUN is defined as the hourly price that equals the total incomes with the total revenues

in the system in each hour, and therefore is strictly linked with the zonal prices. One of the difference in price forecasting between PUN and zonal prices is that zonal events (e.g., a rainy day in the north of Italy while in south it's a sunny day) affect especially the zonal prices, while PUN is not directly affected. Hourly prices are made available by the Italian PEO (Gestore dei Mercati Energetici, GME [10]). There are more than 40 variables that can be used as predictors [1], but we investigated the relation between prices and other variables or information selecting only those that are normally already collected by the PEO, or by TSO. In the following are reported only considerations about PUN. Concerning zonal prices, the relations with the variables are analogous. In the squared brackets are reported the letters we adopted in the study to indicate each predictor.

1. Historical Prices [P]: as historical data set we used the price of the day before for days from Tuesday to Friday, and the price of the week before for the remaining days, since the trend of the price in the week-end is different from the trend during the week.

2. Forecast load [L]: in price forecasting the estimated demand has always played an important role [18] and it seems to be the variable that most influence the price, see Figure 3.2. The hourly data of forecast load have been made available by GME.

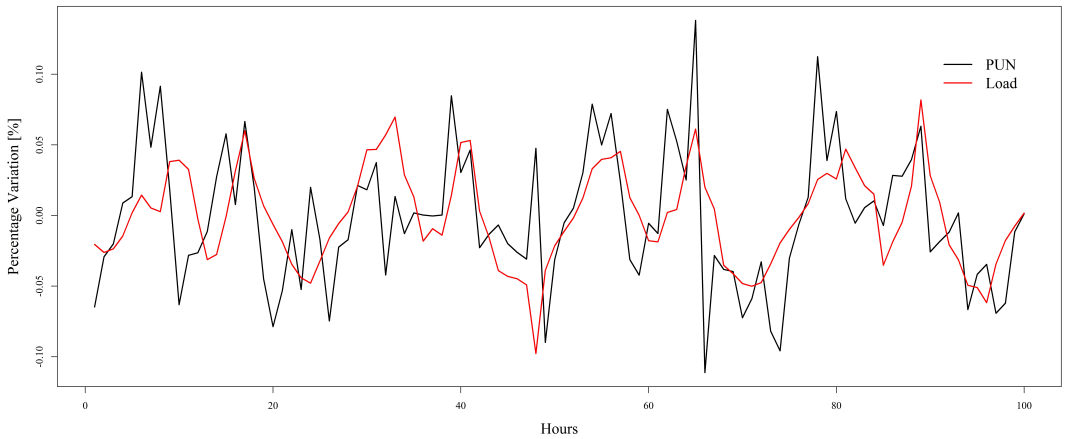


Figure 3.2: *Variation in percentage from one hour to another of PUN (black line) and forecast load (red line)*

3. Wind and Solar Power Production [W]: an high production of renewable energy can decrease the price since it is sold at a price close to zero. For this reason we

used as predictors the hourly wind and solar power forecasts available on the Italian TSO Terna [19] website. Figure 3.3 represents the percentage variation of PUN and wind and solar power production.

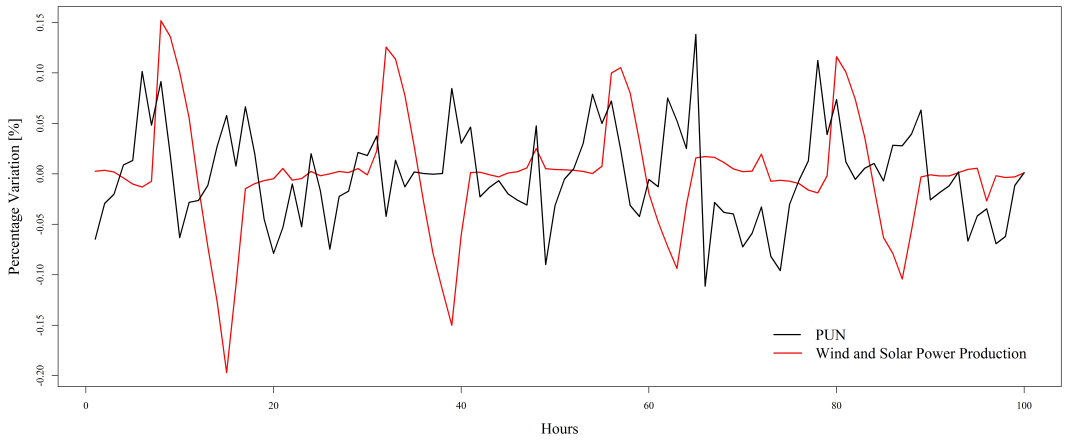


Figure 3.3: *Variation in percentage from one hour to another of PUN (black line) and wind and solar power production (red line)*

4. Plenty or shortage of water [Y]: hydroelectric power production represents around the 48% of the total renewable production [20], but since we did not have forecasted data of hydro power for the day ahead, we considered five levels of weekly production, available on GME website. In Figure 3.4 are reported the bootstrap confidence intervals of each level.

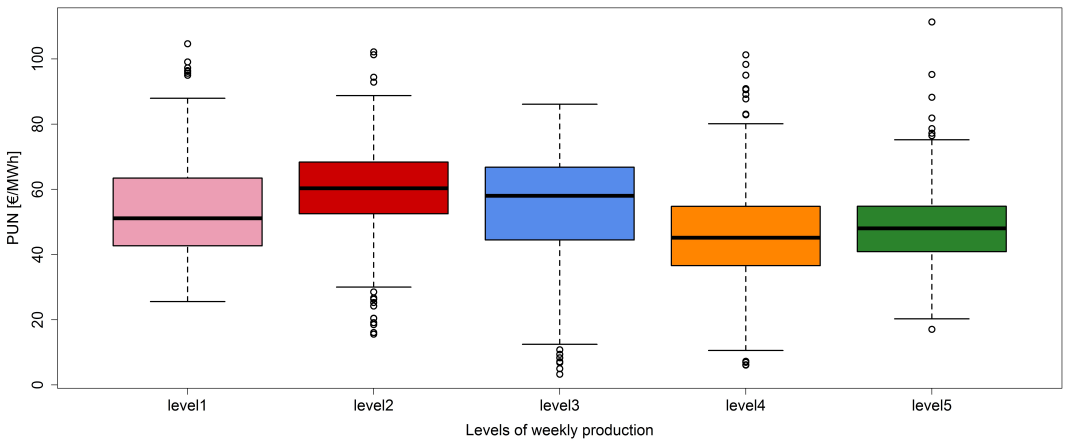


Figure 3.4: *Bootstrap confidence intervals of each level of weekly hydroelectric power production.*

5. The Net Transfer Capacity (NTC) on the national border [N]: Italy imports around the 12% of its requirements, turning out to be the major importer of electricity from abroad in the Europe [19]; hourly data of NTC are available on Terna website.

6. Gas Price [G]: the inflation and the gas prices can affect the electricity price. For this study, only monthly data of gas prices are used, available on a Datastream database.

7. Hour Effects [H]: prices follow particular trends during each day, as shown in Figure 3.1, therefore it is essential to distinguish the hours.

8. Calendar Attributes: events as holidays and festivities affect the price. We characterized each day with number 1 if it is a working day, 2 if it is pre-holiday and 3 if it is holiday [F]. The price data have also some monthly [M] and seasonal [S] variations. In Figure 3.5 are reported the bootstrap confidence intervals of [F], in Figure 3.6 are reported the bootstrap confidence intervals of [M] and in Figure 3.7 are reported the bootstrap confidence intervals of [S].

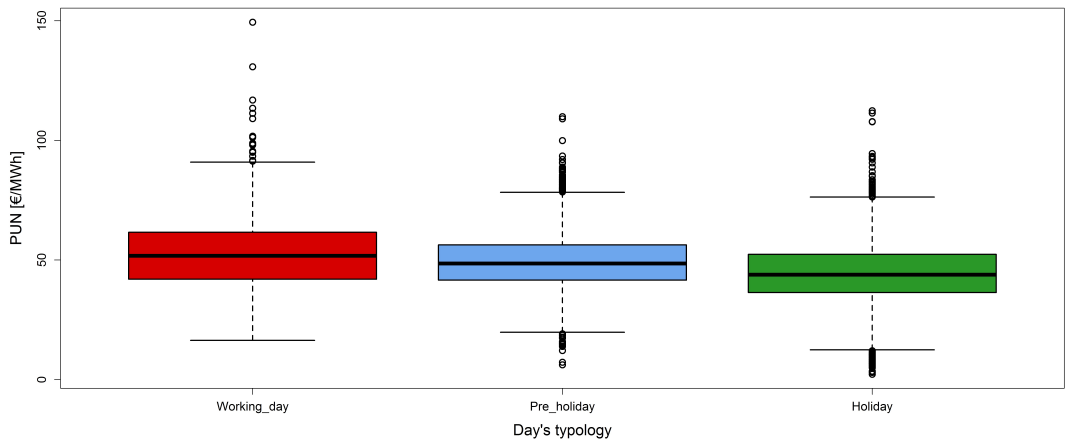


Figure 3.5: *Bootstrap confidence intervals related to festivities.*

Temperature is an important variable that many authors use in price forecasting [1][2], since it is the main exogenous variable that affect the system load. We didn't use this information since we can use directly the forecast load.

We considered national data (i.e. data collected over the entire Italy) to predict the PUN, whereas to predict zonal prices we considered some zonal data (i.e. [P], [W] and

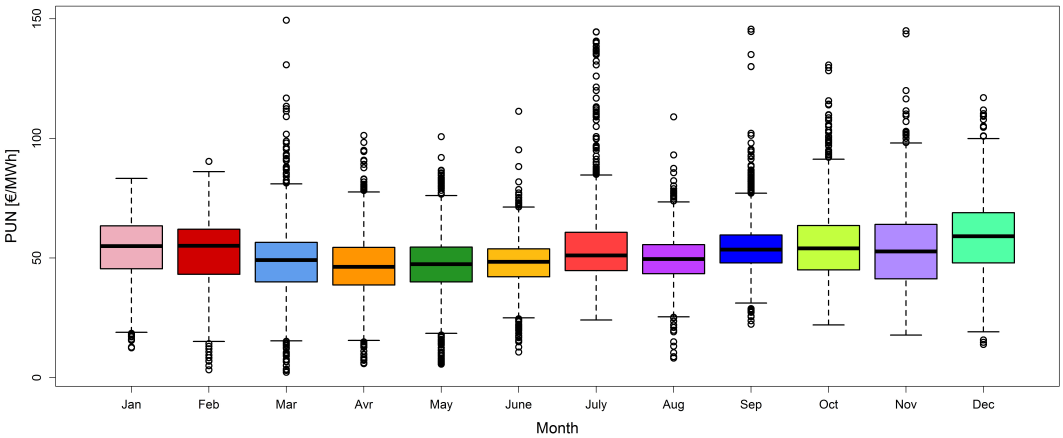


Figure 3.6: *Bootstrap confidence intervals related to months.*

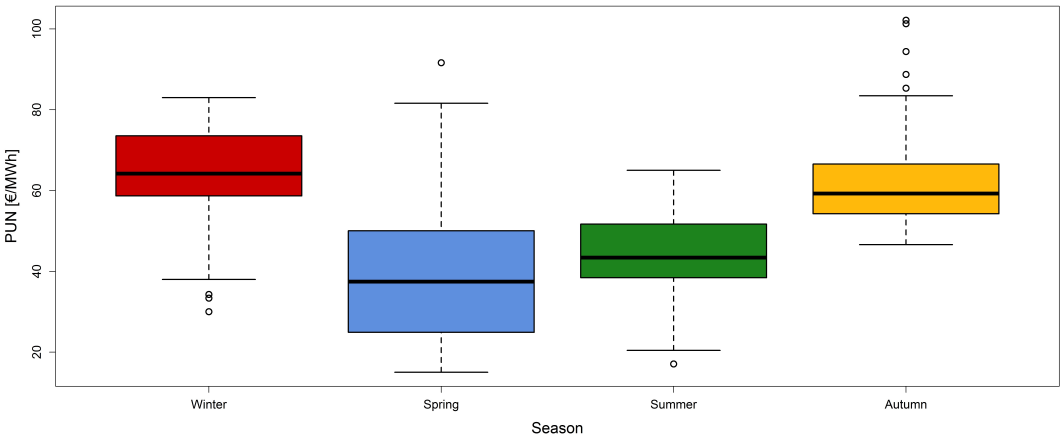


Figure 3.7: *Bootstrap confidence intervals related to seasons.*

[Y] of each zone). Furthermore, the NTC on the national border is kept in consideration only for PUN and Nord price predictions, since the other zones are not adjacent with any countries.

The dataset is split into two parts: a training period of one year (2014) in which NN and SVR are trained, and a test period (2015) that is kept as an independent verification period. We chose one year of training period to cover all the possibilities of calendar attributes.

3.3 Prediction Methods

3.3.1 Neural Network (NN)

An NN is a system of interconnected group of nodes, called neurons, able to link some input variables to one or more output variables. To this end, each neuron assigns a weight to the input signal and with a learning process performed through different iterative cycles it learns how to associate the right output to each input. At each iteration, each input *in* is multiplied by an appropriate signal weight, represented by a value w. This product has as result the value

$$net_i = \sum_{i,j} w_{i,j} \cdot in_j.$$

Given an arbitrary value μ_j, the term $net_i - \mu_j$ represents the argument of the activation function and generates the output. Since the weights are usually in the range $(-1,1)$, the NN input are normalized between 0 and 1, to adapt the activation function to the weights.

The strength of the learning phase is evaluated by the difference between the forecast output signal and the measured value; this difference represents the error. The training period is split into two sub-parts, each one taken randomly. During the first part (that is the 80% of the total training period) an independent iterative training is performed starting from different random weights. The NN that provides the minimum Root Mean Square Error (RMSE), compared to the output over the second part (that therefore is the 20% of the total training period), is selected and applied over the test period. We used a recursive approach, i.e. parts of test period already verified are added to the training dataset, There is not an optimal number of neurons that can be chosen a priori, but in literature it is

possible to find different rules [21][22]. We chose to use $n+1$ neurons where n is the number of input. The back-propagation algorithm used in this work is based on the *nnet* package [23] of the programming language *R* [24].

3.3.2 Support Vector Regression (SVR)

SVR are based on Support Vector Machine (SVM), generally used for data classification and regression analysis. SVR is the methodology by which the input data are mapped into a higher dimensional featured space through a nonlinear mapping ϕ and then a linear regression problem is obtained and solved in this featured space. Give training $(x_1, y_1), \ldots, (x_N, y_N)$ where x_i are input vectors and y_i are the associated output variable of x_i (the price to be estimated), the mapping function, following Vapnik [16], can be formulated as

$$f(x) = \sum_{i=1}^{N} \omega \cdot \phi(x_i) + b$$

where ω and b are parameters to be estimated. The goal is to find the optimal ω and b such that f has at most ε deviations from the output y. The flatness (that we want to minimize) of the weights ω can be measured by the Euclidean norm and therefore we need to solve the following optimization problem:

$$\begin{aligned} &\min_{\omega,b} \frac{1}{2}\|\omega\|^2 \\ &\text{s.t. } \omega^T \cdot \phi(x_i) + b - y_i \leq \varepsilon \\ &\quad\; y_i - \omega^T \cdot \phi(x_i) - b \leq \varepsilon \end{aligned} \tag{3.1}$$

This problem means that any deviation larger then ε is not accepted, while any error less than ε is not take into account. To avoid possible infeasibilities in constraints of optimization problem (3.1), slack variables ξ_i, $\hat{\xi}_i$ are added and the final formulation follows:

$$\begin{aligned} &\min_{\omega,b,\xi_i,\hat{\xi}_i} \frac{1}{2}\|\omega\|^2 + C \cdot \sum_{i=1}^{N} (\xi_i + \hat{\xi}_i) \\ &\text{s.t. } \omega^T \cdot \phi(x_i) + b - y_i \leq \varepsilon + \xi_i \\ &\quad\; y_i - \omega^T \cdot \phi(x_i) - b \leq \varepsilon + \hat{\xi}_i \\ &\quad\; \xi_i, \hat{\xi}_i \geq 0 \end{aligned} \tag{3.2}$$

The parameters that control the regression quality are the cost of error, C, and the width of the tube, ε . This problem is usually solved through its dual, for the following reasons: the matrix associated to the dual problem contains more zero than those of the primal problem and the dual allows to easily working with the nonlinear mapping ϕ. The solution of problem (3.2) is expressed as

$$f(x) = \sum_{i=1}^{N} (\lambda_i - \hat{\lambda}_i) \cdot K(x, x_i) + b$$

where $\lambda_i, \hat{\lambda}_i$ are the dual variables of constraints of problem (3.2) with $0 \leq \lambda_i, \hat{\lambda}_i \leq C$. The Kernel function $K(x_j, x_i) = \phi(x_j) \cdot \phi(x_i)$ satisfies Mercer's conditions [25] and can be of different types: Gaussian, polynomial, exponential, hyperbolic tangent. In this application we chose to use a linear kernel after a sensitivity analysis on the dataset.

Different choices of the parameters are made to find the optimal combination. These combinations are tested on the second part of the training period, as for the NN. Also in this case the SVR that provides the minimum RMSE, compared to the output, is selected and applied over the test period. The algorithm is based on the *e1701* package [26] of the programming language *R*.

3.3.3 Reference Methods

To better understand the contribution of NN and SVR as post-processing techniques, we compare them to two less sophisticated methods, i.e. a linear regression (LR) and the persistency (P) of the data (i.e. considering as forecast the price of the day before for days from Tuesday to Friday and the price of the week before for days from Saturday to Monday). This choice is due to the fact that LR is one of the prediction methods most known and so it is easy to compare. The persistency is the easiest way to provide a prediction, and so a good prediction method must perform highly better. LR is computed through the function *lm* [27] of the programming language *R*.

3.4 Results

3.4.1 Forecast evaluation

RMSE, Mean Absolute Error (MAE), bias and Pearson correlation indices are used to assess the performances of the different prediction systems. Bias is defined as the difference between the arithmetic mean of observed and forecasted values. It is obtained by the formula

$$BIAS = \frac{1}{N} \cdot \sum_{i=1}^{N} (f_i - m_i)$$

where f_i represents the forecasted data and m_i the measured one. The bias evaluates the systematic error and its optimal value is 0; a positive (negative) value indicates a tendency to overestimate (underestimate) the observations.

MAE is given by the formula

$$MAE = \frac{1}{N} \cdot \sum_{i=1}^{N} |f_i - m_i|$$

i.e., it is the average absolute error between the forecasts and the observations. Lower MAE values indicate greater forecast accuracy.

RMSE is the squared root of the residual variance. It is given by the formula

$$RMSE = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^{N} (f_i - m_i)^2}$$

i.e., it is the quadratic average of the errors and it indicates how much the forecasted data are far from the observed data. Similarly to MAE, a lower RMSE indicates greater accuracy.

The Pearson correlation index is a coefficient that assesses the linearity between the covariance of two random variables and the product of their standard deviation, based on the formula

$$cor = \frac{cov(f_i, m_i)}{\sqrt{sd(m_i) sd(f_i)}}$$

The range of this coefficient is $[-1, 1]$. It evaluates how the forecast and the measurement temporal trend are similar. A correlation equal to 1 indicates a perfect forecast.

3.4.2 PUN results

Initially, to understand the importance of each predictor in price forecasting, different simulations have been done adding one by one each predictor. The best result is obtained considering as input all the available predictors, and this holds for all the post-processing methods, even if some singular predictors (e.g. the monthly information) does not seem to be correlated with the price to be predicted. In Table 3.1 are reported the results of the NN applied to different predictors (the results of SVR, LR and P follow the same trend, and therefore they are not reported here).

Table 3.1: *RMSE, MAE, BIAS and correlation of NN applied to different predictors.*

Predictors	**RMSE [€/MWh]**	**MAE [€/MWh]**	**BIAS [€/MWh]**	**Corr.**
[H,P]	7.65	5.20	1.29	0.854
[H,P,F]	7.54	5.26	1.47	0.871
[H,P,F,M]	7.71	5.43	2.11	0.862
[H,P,F,M,L]	7.45	5.35	1.98	0.880
[H,P,F,M,L,Y]	7.35	5.23	1.75	0.873
[H,P,F,M,L,Y,W]	6.97	5.04	1.63	0.876
[H,P,F,M,L,Y,W,G]	6.64	4.78	-0.18	0.884
[H,P,F,M,L,Y,W,G,N]	6.54	4.65	0.015	0.885
[H,P,F,M,L,Y,W,G,N,S]	**6.14**	**4.55**	**0.15**	**0.895**
[H,P,F,L,Y,W,G,N,S]	7.20	4.62	-0.18	0.890

It is possible to observe that, even if adding [M] the results are lower than before (from row 2 to row 3 in Table 3.1), using all the predictors together, the performance with [M] is better than those of without (rows 9 and 10 in Table 3.1). This can be justified considering that a single predictor could not lead any information in the prediction, but could improve the information of the other predictors. For example, information about month can be useful together with the information on solar production, since solar production should increase in hot months.

The four different post-processing methods are compared and evaluated in term of evaluation indices and the results are shown in Table 3.2.

The NN is the post-processing technique that performs better, whereas P is the worst

Table 3.2: *RMSE, MAE, BIAS and correlation of the four prediction methods.* .

	NN	**SVR**	**LR**	**P**
RMSE [€/MWh]	6.142	6.847	7.403	8.704
MAE [€/MWh]	4.554	5.014	5.602	5.683
Bias [€/MWh]	0.152	-0.494	1.065	0.384
Correlation	0.8954	0.8612	0.8488	0.842

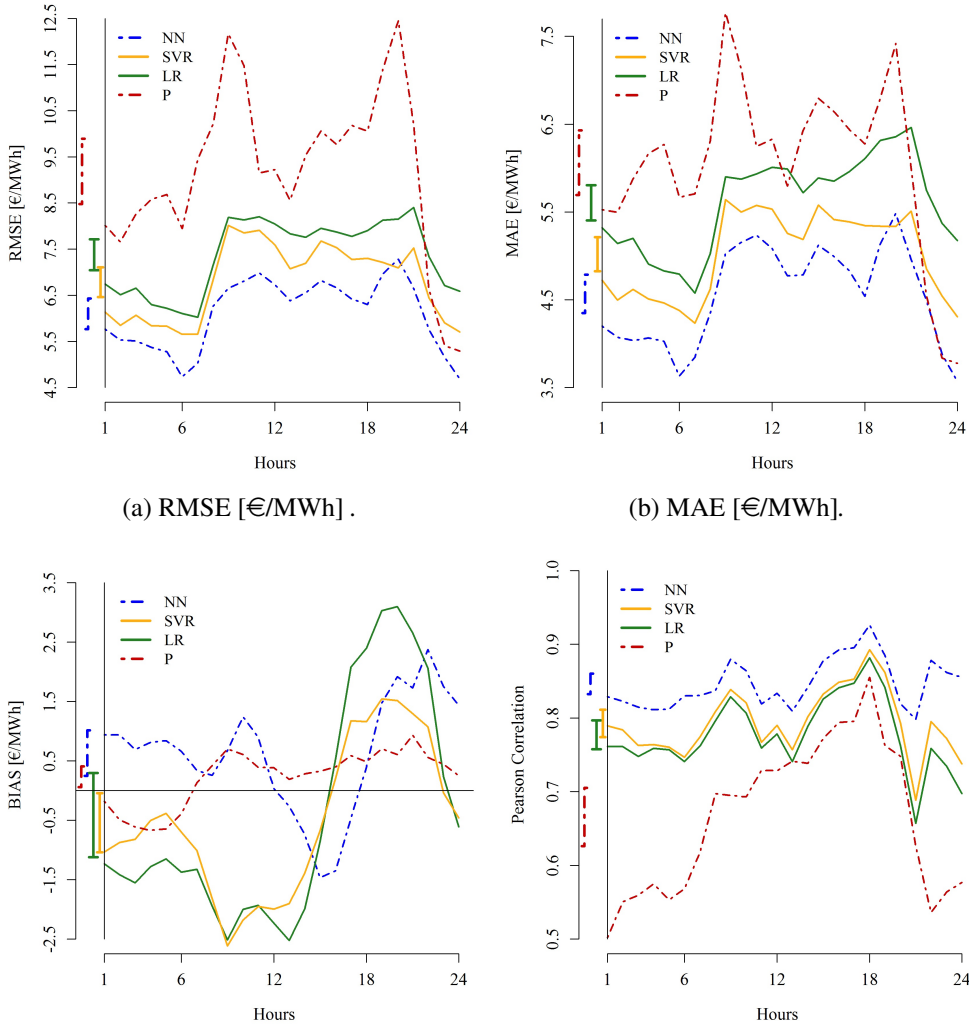
one. The SVR seems to be a good method, but not good as the NN. To better understand how the four methods performs, in Figure 3.2 are shown, respectively, RMSE in €/MWh (a), MAE in €/MWh (b), BIAS in €/MWh (c) and the Pearson correlation (d) of NN, SVR, LR and P.

In all the plots of Figure 3.8, NN performs evidently better than the other methods, and regarding bootstrap confidence intervals in the left of the plots, it is clear that SVR performs better than LR. The trend of NN, SVR and LR are quite similar and they show a greater difficulty in forecast the price in some hour of the day, in particular in the afternoon. This can be justified considering that for example there is more variability in wind speed (hence less predictable conditions) due to stronger turbulent convection in the afternoon hours. Furthermore, solar power is not available during the night, and so in these hours all the solar forecast errors are avoid. Also the request of load is different during the entire day, and therefore the trend of the price forecast denotes the trend of request of load.

The trend of persistency is completely different from the others, and this can mean that historical prices are not sufficient to lead to a good forecast.

3.4.3 Zonal Prices results

The Italian electricity day-ahead market is split into six major geographical zones: North (NORD), Centre-North (CNOR), Centre-South (CSUD), South (SUD), Sicily and Sardinia, where the last two correspond to the major Italian islands. For this study we have only considered the four first zones, and for them the NN, SVR, LR and P have been applied. Differently from PUN, here a local event should evidently affect the forecast. SUD zone, for example, it's very sunny, and therefore the solar and wind production [W]



(a) RMSE [€/MWh] .

(b) MAE [€/MWh].

(c) BIAS [€/MWh].

(d) Correlation.

Figure 3.8: *Indices of NN, SVR, LR and P as a function of the hours. The arrows to the left of each plot represent the bootstrap confidence intervals of each index obtained grouping data over the different hours together.*

should be an important information. As for PUN, also in this case different simulations have been done to understand the importance of a single prediction in the zonal prices forecast. For each zone a different result has been obtained, confirming that the local events are more important it this prediction. In Tables 3.3- 3.4- 3.5- 3.6 are shown RMSE, MAE, bias and correlation of NN, SVR and LR applied to different predictors for the four zones; only the best simulation is reported for each zone and for each method.

Table 3.3: *RMSE, MAE, BIAS and correlation of NN, SVR and LR applied to different predictors for zone **NORD**.*

Method	**Predictors**	**RMSE [€/MWh]**	**MAE [€/MWh]**	**BIAS [€/MWh]**	**Corr.**
NN	**[H,P,F,M,L, Y,W,G,N,S]**	**6.13**	**4.27**	**0.10**	**0.880**
SVR	[H,P,F,M,L, Y,W,G,N,S]	7.15	4.90	-1.74	0.87
LR	[H,P,F,M, L,Y,W, G]	7.18	5.13	-0.46	0.722

Table 3.4: *RMSE, MAE, BIAS and correlation of NN, SVR and LR applied to different predictors for zone **CNOR**.*

Method	**Predictors**	**RMSE [€/MWh]**	**MAE [€/MWh]**	**BIAS [€/MWh]**	**Corr.**
NN	**[H,P,F,L, Y,W,G]**	**8.04**	**5.46**	**-0.51**	**0.839**
SVR	[H,P,F,L, Y,W,G]	8.55	5.68	-1.28	0.838
LR	[H,P,F, L,W]	8.94	6.16	1.28	0.699

The predictors that lead the best forecast change both for each zone and for each method. For example, LR needs less predictors than the NN and SVR. This can be explained considering that LR is a linear regression, while NN and SVR are able to work also with non-linearity between input and output. As consequence, NN and SVR perform better than LR and they can explain more predictors. In NORD the best forecast is obtained using the same predictors as for PUN.

This holds also for SUD, even if predictor [N] misses since SUD is not adjacent with any countries. Concerning CNOR and CSUD, the best prediction is obtained using all the predictors less the information about months and seasons.

Table 3.5: *RMSE, MAE, BIAS and correlation of NN, SVR and LR applied to different predictors for zone **CSUD**.*

Method	Predictors	RMSE [€/MWh]	MAE [€/MWh]	BIAS [€/MWh]	Corr.
NN	**[H,P,F,L, Y,W,G]**	**8.04**	**5.65**	**-0.66**	**0.828**
SVR	[H,P,F,L, Y,W,G]	9.27	6.07	-1.53	0.819
LR	[H,P,F, L,W]	9.42	6.28	1.19	0.667

Table 3.6: *RMSE, MAE, BIAS and correlation of NN, SVR and LR applied to different predictors for zone **SUD**.*

Method	Predictors	RMSE [€/MWh]	MAE [€/MWh]	BIAS [€/MWh]	Corr.
NN	**[H,P,F,M,L, Y,W,G,S]**	**8.98**	**6.27**	**0.09**	**0.823**
SVR	[H,P,F, L,Y,W]	9.22	6.72	1.74	0.810
LR	[H,P,F,L]	9.83	6.98	1.06	0.59

For all the four zones, the best forecast is obtained using the NN, and SVR indices are quite similar to those of LR. Regarding NORD, all the three methods seem to perform as for the PUN, while in the other zones it seems that it is more difficult to obtain a good forecast. This can be due to zonal effects that in PUN case are deadened by the other zones, while here are difficult to predict. One of the reasons such that NORD is the easiest zone to forecast, is that it is adjacent with other countries, and so its price strictly depends from those of France, Switzerland, Austria and Slovenia.

In Table 3.7 are summarized the best combinations of predictors for each zone.

Figure 3.9 shows index MAE as a function of hours for the four zone obtained with the best method, i.e. NN. It is evident how in NORD is easiest to forecast the zonal price. The

Table 3.7: *Best combination of predictors for each zone.*

	[P]	[H]	[L]	[W]	[Y]	[G]	[N]	[F]	[M]	[S]
NORD	X	X	X	X	X	X	X	X	X	X
CNOR	X	X	X	X	X	X		X		
CSUD	X	X	X	X	X	X		X		
SUD	X	X	X	X	X	X		X	X	X

trend of CNOR, CSUD and SUD are similar, that is the price in these zones is conditioned by the same variables. As for the PUN, the hours in the afternoon are more difficult to forecast. RMSE, bias and correlation (here not reported) show similar trends.

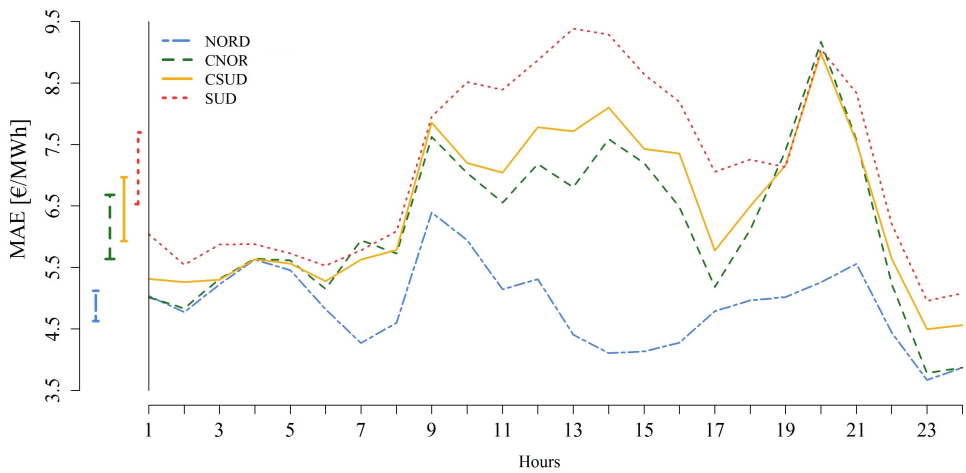


Figure 3.9: *MAE [€/MWh] of NORD, CNOR, CSUD and SUD, obtained with NN, as a function of the hours. The arrows to the left of each plot represent the bootstrap confidence intervals of each index obtained grouping data over the different hours together.*

3.5 Conclusion

We tested different approaches and the importance of different predictors to forecast the hourly Italian prices, that are the uniform purchase price (Prezzo Unico Nazionale, PUN), that is the same for most of the Italian demand orders, and the zonal prices, that are equal in case there are no congestion, otherwise they depend on the events in each zone.

We applied a Neural Network (NN) and a Support Vector Regression (SVR) to generate predictions of the prices. This methods are been applied on different combinations of predictors, to understand and verify the importance of each predictor in price forecast both for PUN and zonal prices. In both PUN and zonal prices NN leads to better results than those obtained by using SVR. The two methods have been compared with two less sophisticate methods, a simple bias reduction as linear regression (LR) and a Persistency (P) (i.e. using as output the historical prices), and both the two last methods performs worst than NN and SVR.

Regarding the predictors, it seems that, even if a single predictor does not lead a benefit in the forecast, all the predictors taken together lead to the best forecast. This is particularly true for PUN, NORD and SUD price, while for CNOR and CSUD the best configuration is those with all the available predictors less the information about months and seasons.

We have used only data available on different websites, to keep the methodology actually usable by each market participant.

Notes

This chapter has been published as article for the AEIT International Annual Conference, Capri, 5-7 October 2016.

The co-authors are Alberto Gelmini (RSE), Maria Teresa Vespucci (University of Bergamo), Paolo Grisi (RSE), Dario Ronzio (RSE).

References

[1] S. K. Aggarwal, L. M. Saini and A. Kumar. *Electricity price forecasting in deregulated markets: a review and evaluation*. International Journal of Electrical Power and Energy System, vol. 31, n. 1, pp. 13-22, 2009.

[2] R. Weron. *Electricity price forecasting: a review of the state-of-the-art with a look into the future*. International Journal of Forecasting, vol. 30, n. 4, pp. 1030-1081, 2014.

[3] A. D. Papalexopoulos and T. C. Hesterberg. *A regression-based approach to short-term system load forecasting*. IEEE Transaction on Power System, vol. 5, n. 4, pp. 1535-1547, 1990.

[4] A. Kian and A. Keyhani. *Stochastic price modelling of electricity in deregulated energy markets*. Proceedings of the 34th annual Hawaii International Conference on System Science, pp. 1-7, 2001.

[5] J. C. Cuaresma, J. Hlouskova, S. Kossmeier and M. Obersteiner. *Forecasting electricity spot-prices using linear univariate time-series models*. Applied Energy, vol. 77, n. 1, pp. 87-106, 2004.

[6] A. Troncoso Lora, J. M. Riquelme Santos, A. Gomez Exposito and J. M. Martynez Ramos. *Electricity market price forecasting based on weighted nearest neighbours techniques*. IEEE Transaction on Power System, vol. 22, n. 3, pp. 1294-1301, 2007.

[7] H. A. Nielsen, H. Madsen and T. S. Nielsen. *Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts*. Wind Energy, vol. 9, pp. 95-108, 2006.

[8] K. Maciejowska, J. Nowotarski and R. Weron. *Probabilistic forecasting of electricity spot prices using factor quantile regression averaging*. International Journal of Forecasting, vol. 32, n. 3, pp. 957-965, 2016.

[9] T. Senjyu, P. Mandal, K. Uezato and T. Funabashi. *Next day load curve forecasting using hybrid correction method*. IEEE Transaction on Power System, vol. 20, n. 1, pp. 102-109, 2005.

[10] GME. *Vademecum della borsa Elettrica [Power exchange vademecum]*, 2012. https://www.mercatoelettrico.org/it/MenuBiblioteca/Documenti/20091028VademecumBorsaElettrica.pdf.

[11] P. Mandal, T. Senjyu and T. Funabashi. *Neural networks approach to forecast several hour ahead electricity prices and loads in deregulated market*. Energy Conversion Manage, vol. 47, pp. 2128–2142, 2006.

[12] N. Amjady. *Day-ahead price forecasting of electricity markets by a new fuzzy neural network*. IEEE Transaction on Power System, vol. 21, n. 2, pp. 887-896, 2006.

[13] B. Chen, M. Chang and C. Lin. *Load forecasting using support vector machines: a study on EUNITE competition 2001*. IEEE Transaction on Power System, vol. 19, n. 4, pp. 1821-1830, 2001.

[14] C. J. C. Burges. *A tutorial on support vector machines for pattern recognition*. Data Mining and Knowledge Discovery, vol. 2, pp. 121-167, 1998.

[15] C. Gao, E. Bompard, R. Napoli and H. Cheng. *Price forecast in the competitive electricity market by support vector machine*. Physica A: Statistical Mechanics and its application, vol. 382, n. 1 pp. 98-113, 2007.

[16] V. N. Vapnik. *The nature of statistical learning theory*. Springer, NY, 1995.

[17] V. N. Vapnik. *Statistical learning theory*. Wiley, NY, 1998.

[18] P. S. Georgilakis. *Market clearing price forecasting in deregulated electricity markets using adaptively trained neural networks*. Lecture notes in Advances in Artificial Intelligence, vol. 3955, pp. 56-66, 2006.

[19] Terna. http://www.terna.it/.

[20] GSE. *Rapporto statistico, energia da fonti rinnovabili, anno 2014 [Statistical report, energy from renewable power plants, year 2014]*. http://www.gse.it/it/salastampa/GSE_Documenti/Rapporto%20statistico%20GSE%20 -%202014.pdf.

[21] Z. Boger and H. Guterman. *Knowledge extraction from artificial neural network models*. IEEE Systems, Man and Cybernetics Conference, Orlando, FL, USA, vol. 4, pp. 3030-3035, 1997.

[22] A. Blum. *Neural Network in C++*. Wiley, NY, 1992.

[23] W. N. Venables and B. D. Ripley. *Modern applied statistics with S*. Fourth edition, Springer, NY, ISBN 0-387-95457-0, 2002.

[24] *R, a language and environment for statistical computing*. R foundation For Statistical Computing, Vienna, Austria, ISBN: 3-900051-07-0.

[25] N. I. Sapankevych and R. Sankar. *Time series prediction using support vector machine: a survey*. IEEE Computational Intelligence Magazine, vol. 4, pp. 24-38, 2009.

[26] D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel and F. Leisch. *e1071*. Misc functions of the department of statistics, probability theory group, TU, Wien, 2015.

[27] J. M. Chambers. *Linear models*. Chapter 4 of Statistical models in S, eds J. M. Chambers and T.J. Hastie, Wadsworth & Brooks/Cole, 1992.