4.1.1 Very-Short–Term Forecasting

4.1.1 Very-Short–Term Forecasting

4.1.1.1 Wind Speed Forecasting Using Statistical Methods

This approach has to do with wind speed forecasting only. However, this wind speed forecast is often converted to power through an empirical or manufacturer’ s power curve.

这种方法只与风速预测有关。 但是，这个风速预测通常通过经验或制造商的功率曲线转换为功率。

The first wind forecasting model, specifically used in wind generation forecasting, is presented in [90]. A Kalman filter [91] that uses the last six measured values as inputs is proposed to forecast wind speed for the following minutes. The results are good when compared with the persistence for time horizons below 10 min. of averaged data. The improvement was poorer in longer averages and was nonexistent for 1-hr averages. Also, in [92], a Kalman filter is used to control a variable speed wind turbine.

第一个风力预报模型，具体用于风力发电预报，在[90]。 使用最后六个测量值作为输入的卡尔曼滤波器[91]来预测接下来的分钟的风速。 当与10分钟以下时间范围的持续性相比时，结果是好的。 的平均数据。 改善在较长的平均值中较差，并且对于1小时平均值不存在。 此外，在[92]中，卡尔曼滤波器用于控制变速风力涡轮机。

For models based on the Box-Jenkins methodology, the general Autoregressive Integrated Moving Average (ARIMA) approach [93] was the first to be employed on wind speed forecasting. For example, Contaxis et al. [94] employed an autoregressive (AR) model (more precisely an AR[3]) to forecast the wind speed for time horizons ranging between 30 min. and 5 hr and used the values to control an isolated hybrid diesel/wind system and short-term operation scheduling; Kamal et al. [95] used an ARIMA model to forecast the wind speed and estimate confidence intervals; Schlink et al. [96] employed these models to forecast the wind speed for the next 10 minutes in an airport; Poggi et al. [97] used an autoregressive model for each month in order to forecast the wind speed for the following 3 hr; Torres et al. [98] used five Auto-Regressive Moving Average (ARMA) models to forecast the hourly average wind speed for a time horizon of 10 hr in five different locations with different topographic characteristics. With this model, over nine years it was possible to achieve a 20% error reduction as compared to persistence; Tantareanu [99] found that ARMA models can perform up to 30% better than persistence for 3 to 10 steps ahead in 4-s averages of 2.5-Hz sampled data.

对于基于Box-Jenkins方法的模型，一般的自回归差分移动平均（ARIMA）方法[93]是第一个应用于风速预测的方法。例如，Contaxis et al。 [94]采用自回归（AR）模型（更精确地说是AR [3]）来预测30分钟之间的时间层的风速。和5小时，并使用该值来控制孤立的混合动力柴油/风力系统和短期操作调度; Kamal et al。 [95]使用ARIMA模型预测风速和估计置信区间; Schlink等人[96]采用这些模型来预测机场中接下来10分钟的风速; Poggi et al。 [97]每月使用自回归模型，以便预测以下3小时的风速; Torres et al。 [98]使用五个自回归移动平均（ARMA）模型来预测在具有不同地形特征的五个不同位置中的10小时时间范围内的小时平均风速。使用这个模型，九年以来，可以实现与持久性相比20%的误差减少; Tantareanu [99]发现，在2.5 Hz采样数据的4 s平均值中，ARMA模型可以比持续3至10步提高30%。

Kavasseri et al. [100] presented the fractional-ARIMA (f-ARIMA) model to forecast the wind speed day-ahead and two-days-ahead. The forecasted wind speed is converted to wind power by using a manufacturer’ s power curve. Following the results presented in [101], the authors suggest a modified ARIMA model (f-ARIMA) to deal with long-range correlations (LRCs). An LRC process is characterized by a slow decay of the autocorrelation function. The f-ARIMA model allows the differencing parameter to assume fractionally continuous values in the interval [–0.5;0.5], and therefore the differencing parameter can represent LRC.

Kavasseri et al。 [100]提出了分数-ARIMA（f-ARIMA）模型来预测前后和前两天的风速。 预测的风速通过使用制造商的功率曲线转换为风力。 根据[101]中提出的结果，作者建议修改的ARIMA模型（f-ARIMA）来处理长程相关性（LRC）。 LRC过程的特征在于自相关函数的缓慢衰减。 f-ARIMA模型允许差分参数在区间[-0.5; 0.5]中采取分数连续的值，因此差分参数可以表示LRC。

El-Fouly et al. [102] presents a new technique to forecast wind speed for the upcoming hour based on the Grey predictor model [103]. The wind speed is then converted to wind power by a manufacturer’ s power curve. The registered improvement in comparison to persistence is in the range of 12% for the wind generation forecast.

El-Fouly et al。 [102]提出了一种基于灰色预测模型预测即将到来的小时的风速的新技术[103]。 然后通过制造商的功率曲线将风速转换成风力。 与持续性相比，记录的改进在风力发电预报的12%的范围内。

Alternative forecast methods are based on artificial intelligence techniques, namely neural networks (NNs), support vector machines (SVMs), or fuzzy inference systems (FISs). Alexiadis et al. [104] proposed a model based on an NN to forecast the wind speed for the Syros island (in Greece) using historical wind data from the island and from other neighboring islands as input variables. The results show an improvement of 32% over persistence in the forecast error for a 1-hr horizon; the same method was employed in [105] for a different location in Greece, and the improvement over persistence was 27% for a 2-hr time horizon.

备选的预测方法基于人工智能技术，即神经网络（NN），支持向量机（SVM）或模糊推理系统（FIS）。 Alexiadis et al。 [104]提出了一个基于NN的模型，以使用来自岛屿和其他相邻岛屿的历史风力数据作为输入变量来预测Syros岛（希腊）的风速。 结果表明，对于1小时水平的预测误差，持续性改善了32%; 在[105]中为希腊的不同位置采用相同的方法，并且在2小时时间范围内对持续性的改善为27%。

Sfetsos evaluates different models in [106]: a persistence model, ARIMA models, NN, and neuro-fuzzy systems. The model with the best results was the NN, leading to a 20–40% average improvement when compared to the persistance. In more recent work, Sfetsos [107] uses two models based on NN to forecast the wind speed for a time horizon of one hour. The first model uses the last known values of the hourly wind speed as inputs, and the results are only 3% better than those registered for persistence models. The second model uses the wind speed time series with 10-min. intervals as inputs, in addition to using the NN output iteratively to forecast the subsequent 60 min. The improvement of the second model over persistence is 10%.

fetsos评估[106]中的不同模型：持久性模型，ARIMA模型，NN和神经模糊系统。 具有最佳结果的模型是NN，与持久性相比，导致20-40%的平均改善。 在最近的工作中，Sfetsos [107]使用两个基于NN的模型来预测一个小时的时间范围内的风速。 第一个模型使用每小时风速的最后已知值作为输入，结果仅比持续性模型注册的结果好3%。 第二个模型使用10分钟的风速时间序列。 间隔作为输入，除了使用NN输出迭代地预测随后的60分钟。 第二个模型对持续性的改进是10%。

Cadenas and Rivera [108] tested four NN configurations to hourly wind speed forecast.
The model with best performance was the simple one, an NN with two layers and three neurons.

Cadenas和Rivera [108]对每小时的风速预报测试了四种NN配置。具有最佳性能的模型是简单的，具有两层和三个神经元的NN。

Damousis and Dokopoulos [109] and Damousis et al. [110] present a Takagi-Sugeno FIS [111] that is based on wind measures of the target location and on the wind speed forecasts of neighboring locations for a time horizon of between 30 and 240 min. A genetic algorithm is used in order to optimize the FIS parameters. The improvement over persistence ranges between 9.5% and 28.4%, depending on the time horizon (it increases with the time horizon).

Damousis和Dokopoulos [109]和Damousis et al。 [110]提出了基于目标位置的风力测量以及在30和240分钟之间的时间范围上的相邻位置的风速预测的Takagi-Sugeno FIS [111]。 使用遗传算法以优化FIS参数。 持续性的改进范围在9.5%和28.4%之间，取决于时间范围（随时间范围增加）。

Maqsood et al. [112] present the idea of using more than one model to forecast three meteorological variables (including wind speed) for a 24-hr-ahead interval. Four different types of NNs [113] are considered: the multilayer perceptron (MLP), the recurrent neural network of Elman, the radial basis function (RBF), and the Hopfield neural networks. An NN of each type was trained for each season of the year. The best result was the one obtained with the RBF neural network, but accuracy increases when all of the models are combined (i.e., into an ensemble of models).

Maqsood et al。 [112]提出了使用多个模型来预测24小时前的时间间隔的三个气象变量（包括风速）的想法。 四种不同类型的NNs [113]被考虑：多层感知器（MLP），Elman的复现神经网络，径向基函数（RBF）和Hopfield神经网络。 每年的每个季节训练每种类型的NN。 最好的结果是用RBF神经网络获得的结果，但是当所有模型组合时（即，到模型的整体），精度增加。

Abdel-Aal et al. [114] demonstrated the application of abductive networks based on the group method of data handling (GMDH) [115] to forecast the mean hourly wind speed. The authors demonstrated that the main advantage of the abductive networks over the NN is the fast convergence during training and automatic selection of both input variables and model structure. The model achieved an improvement of 8.2% compared to persistence in a 1-hr-ahead forecast. The model was also used to forecast the wind speed for 6 and 24 hr ahead, achieving an improvement over persistence of 14.6% and 13.7%, respectively.

Abdel-Aal et al。 [114]演示了基于数据处理组（GMDH）[115]的预测平均小时风速的诱导网络的应用。 作者证明，诱发网络相对于NN的主要优点是训练期间的快速收敛和输入变量和模型结构的自动选择。 与1小时前预测中的持续性相比，该模型实现了8.2%的改进。 该模型还用于预测6和24小时的风速提前，实现持续性的改善分别为14.6%和13.7%。

Potter [116] presents an Adaptive Neural Fuzzy Inference System (ANFIS) [117] to forecast the wind speed for a 2.5-min. time horizon. The input data is the measured wind speed, which is then adjusted through splines that considerably decrease the forecast error relative to persistence.

Potter [116]提出了一个自适应神经模糊推理系统（ANFIS）[117]来预测风速2.5分钟。 时间范围。 输入数据是测量的风速，然后通过样条进行调整，这显着地减少了相对于持续性的预测误差。

Ramírez-Rosado and Fernández-Jiménez [118] and [119] presented a three-phase model: (i) the Fourier transform of the last 24 values of mean wind speed is computed; (ii) 23 fuzzy inference systems (Takagi-Sugeno) forecast the coefficients of the Fourier transform for the following hour; and (iii) the mean wind speed is forecasted for the following hour based on the forecasted coefficients of the Fourier transform.

Ramírez-Rosado和Fernández-Jiménez[118]和[119]提出了一个三相模型：（i）计算最后24个平均风速值的傅立叶变换;（ii）23个模糊推理系统（Takagi-Sugeno）预测随后小时的傅里叶变换的系数; 和（iii）基于傅立叶变换的预测系数，预测下一小时的平均风速。

A different approach consists of forecasts in the frequency domain. Two works regarding wind speed forecasting use Discrete Hilbert Transform (DHT) can be found in [120] and [121].

不同的方法包括频域中的预测。 关于风速预测使用的两个作品离散希尔伯特变换（DHT）可以在[120]和[121]中找到。

4.1.1.2 Wind Power Forecasting

Another possibility for very short-term forecasting is to forecast wind generation directly, without a previous step in which the wind speed is forecasted.

非常短期预测的另一种可能性是直接预测风力发电，而没有预测风速的先前步骤。

Kariniotakis et al. [122] and [123] tested various forecasting methods for the Greek island of Crete: adaptive linear models, adaptive fuzzy logic models, and wavelet-based models. Adaptive-fuzzy-logic–based models were installed for on-line operation in the context of the Joule II project CARE.

Kariniotakis et al。 [122]和[123]测试了希腊克里特岛的各种预测方法：自适应线性模型，自适应模糊逻辑模型和基于小波的模型。在Joule II项目CARE的背景下安装了基于自适应模糊逻辑的模型用于在线操作。

Ramírez-Rosado and Fernández-Jiménez [124] employed fuzzy time series to forecast the wind generation for a time horizon of 24 hr. Fuzzy time series were coupled with fuzzy linguistic information about wind, such as "strong wind" (e.g., given by an expert), which allowed the forecasting method to register an improvement of 14.3% over persistence. The same authors [125] presented a model based on grouping historical data by using a substractive clustering method [126]. For each group, a linear regression model is employed to forecast wind generation. The forecasted value is the weighted mean of all values obtained by each group's regression model. The time horizon is 6 hr, and the improvement over persistence in this horizon was around 14%.

Ramírez-Rosado和Fernández-Jiménez[124]采用模糊时间序列来预测24小时时间范围内的风力发电。 模糊时间系列与关于风的模糊语言信息（例如"强风"（例如，由专家给出））耦合，这允许预测方法登记对持续性的14.3%的改进。 同样的作者[125]提出了一个基于分组历史数据的模型，通过使用一个消极的聚类方法[126]。 对于每个组，使用线性回归模型来预测风力发电。 预测值是通过每个组的回归模型获得的所有值的加权平均值。 时间范围是6小时，在这个地平线上持续性的改善约为14%。

Frías et al. [127] developed a model with an intention to participate in the Spanish intradaily energy markets. The model was based on ANFIS and uses online generation data of wind farms jointly with forecasts for the daily market. The model focuses on short forecasting 48 horizons of up to 10 hr ahead. In order to find the best ANFIS architecture, a heuristic method that combines quantity and type of membership functions of the input variables was used, optimizing operative time through a selection of training samples.

Fríaset al。 [127]开发了一个模型，意图参加西班牙的日内能源市场。 该模型基于ANFIS，并使用风电场的在线生成数据与每日市场的预测。 该模型的重点是短期预测高达10小时前48个视野。 为了找到最好的ANFIS体系结构，使用组合输入变量的隶属函数的数量和类型的启发式方法，通过选择训练样本来优化操作时间。

The California Independent System Operator (CAISO) prototype forecasting algorithm for short-term forecasting is described in [128]. A modified ARIMA model is used to compute the 2.5-hr ahead forecasted growth/decline factor. The model coefficients are adapted on-line, and a bias self-compensation scheme was included in the model with the introduction of an additional term into the modified ARIMA model. The model presents goods results in the first two hours, where the MAE is below 3% and 8% respectively of the maximal observed generation. The authors stressed the need to include NWP and unit status information into the model. Nevertheless, Milligan et al. [129] carried on research to understand to what extent timeseries analysis can improve simple persistence forecasts, as well their usefulness in hour-ahead markets. The ARMA models for both wind speed and wind power output are tested with different parameters. The authors concluded that the capacity of ARMA forecast models differed when applied to different time periods. The authors suggest the possibility of using an ensemble of models instead of a single model.

加利福尼亚独立系统运营商（CAISO）用于短期预测的原型预测算法在[128]中描述。使用修改的ARIMA模型来计算2.5小时前向预测的增长/下降因子。模型系数在线适应，并且偏差自补偿方案包括在模型中，在修改的ARIMA模型中引入了附加项。该模型在前两个小时呈现商品结果，其中MAE分别低于最大观察世代的3%和8%。作者强调需要在模型中包括NWP和单元状态信息。然而，Milligan等人[129]进行研究以了解时间分析可以在多大程度上改善简单的持续性预测，以及它们在时间提前市场中的有用性。用于风速和风力功率输出的ARMA模型用不同的参数测试。作者得出结论，ARMA预报模型的能力在应用于不同时期时有所不同。作者建议使用模型的整体而不是单个模型的可能性

The use of an AR model in WPF was analyzed by Duran et al. in [130]. The authors carried out several tests to select the AR order. Consequently, the authors stated that the order does not depend on the training period. Rather, the optimal order depends on the wind farm (e.g., terrain complexity) and the time horizon of the forecast. The best model found by the authors was an AR of order 11. However, from our experience with Portuguese wind farms, this order is very high, and it is even hard to find AR with an order above 2 in the WPF literature. The improvement over persistence in three wind farms ranges between 3% and 17%. The standard deviation of the error is also lower in the AR when compared with persistence. The results obtained for the independent and aggregated wind farms (the sum of the three independent wind farms) have shown that the aggregation reduces uncertainty and forecast error. For instance, for a time horizon of 6 hr, the aggregation leads to a 23.1% improvement.

Duran等人分析了WPF中AR模型的使用。 在[130]。 作者进行了几个测试来选择AR命令。 因此，作者说，该顺序不依赖于训练期。 相反，最佳顺序取决于风电场（例如，地形复杂性）和预测的时间范围。 作者发现的最好的模型是秩序11的AR。然而，根据我们对葡萄牙风力发电场的经验，这个顺序非常高，在WPF文献中甚至很难找到高于2的AR。 三个风电场持续性的改善在3%和17%之间。 与持续性相比，AR中误差的标准偏差也较低。 对于独立和聚集的风电场（三个独立的风电场的总和）获得的结果表明，聚合减少不确定性和预测误差。 例如，对于6小时的时间范围，聚集导致23.1%的改进。

Costa et al. [131] tested a purely and fuzzy autoregressive, as well as an MLP NN, in order to forecast for 10 steps ahead with 10-min. time steps. The only inputs were measured time series. The models were tested in three wind farms located in Spain. The neural network reached the best overall performance.

Costa et al。 [131]测试了一个纯模糊的自回归，以及一个MLP NN，以便预测10步前10分钟。 时间步长。 唯一的输入是测量的时间序列。 这些模型在位于西班牙的三个风力发电场进行测试。 神经网络达到最佳的整体性能。

Kusiak et al. in [132] tested five different data-mining models [133] [134] to forecast the wind power: SVM, MLP NN, the M5P tree algorithm, the Reduced Error Pruning tree, and the bagging tree. The SVM and MLP NN performed particularly well. The SVM provided accurate forecasts from 10 min. up to 1 hr, while the MLP NN is accurate in forecasts of up to 4 hr.

Kusiak et al。 在[132]测试了五个不同的数据挖掘模型[133] [134]来预测风力：SVM，MLP NN，M5P树算法，减少错误修剪树和装袋树。 SVM和MLP NN表现特别好。 SVM提供了精确的预测，从10分钟。 高达1小时，而MLP NN在高达4小时的预测中是准确的。

Onshore generation yields smoothing power fluctuations because the wind farms are usually spread over a large area [82]. This smoothing effect in the offshore generation is uncommon because the wind turbines are concentrated in a single location and, therefore, the power fluctuations can reach significant levels. The modeling of the offshore fluctuations is currently a challenge in WPF. A discussion of these aspects is available in [135].

陆上发电产生平滑功率波动，因为风电场通常分布在大面积上[82]。 在海上发电中的这种平滑效应是不常见的，因为风力涡轮机集中在单个位置，因此，功率波动可以达到显着的水平。 离岸波动的建模目前是WPF的一个挑战。 这些方面的讨论可参见[135]。

Pinson et al. [136] reported the use of statistical regime-switching models for situations with successive periods of fluctuations with large and lower magnitude. Three types of models are discussed and presented by the authors: the self-exciting threshold autoregressive (SETAR), the smooth transition autoregressive (STAR), and the Markov-switching autoregressive (MSAR). The performance of the models was evaluated on a one-step-ahead forecast (10 and 15 min.) in two Danish wind farms and afterwards compared with the ARMA linear model. In all test cases, the MSAR models significantly outperformed the other models. There is also a gain in applying the SETAR and STAR models instead of ARMA, although it is not significant. The authors concluded that the MSAR captures the influence of some complex meteorological variables on the power fluctuations. They also demonstrate that the regime sequence leading successive periods with different behaviors is very complex and cannot be considered as a simple function of the wind generation level.

Pinson等人[136]报告了统计制度转换模型对于具有大幅度和更小幅度的连续波动周期的情况的使用。三种类型的模型由作者讨论和提出：自励阈值自回归（SETAR），平滑过渡自回归（STAR）和马尔科夫切换自回归（MSAR）。模型的性能在两个丹麦风电场的一个向前预测（10和15分钟）上进行评估，然后与ARMA线性模型进行比较。在所有测试案例中，MSAR模型显着优于其他模型。在应用SETAR和STAR模型而不是ARMA时也有一个增益，尽管它不重要。作者得出结论，澳门特别行政区捕获一些复杂的气象变量对功率波动的影响。他们还表明，导致具有不同行为的连续周期的状态序列是非常复杂的，并且不能被认为是风生成水平的简单函数。

Pinson et al. [137] and Pinson and Madsen [138] improved the previously described MSAR model. A time-variant estimation of the model coefficients is described, as well as a regularization term that enables the reduction of the variability of the model coefficients' estimates. In addition, predictive densities are provided by a combination of conditional densities in each regime. Their quantiles can then be computed by numerical integration methods.

Pinson等人 [137]和Pinson和Madsen [138]改进了以前描述的MSAR模型。 描述模型系数的时变估计以及使得能够减少模型系数估计的变异性的正则化项。 此外，通过每个方案中的条件密度的组合提供预测性密度。 然后可以通过数值积分法计算它们的分位数。

4.1.2 Short-Term Forecasting Using NWP

4.1.2.1 Statistical and Computational Intelligence Techniques Applied to WPF

In the literature, several techniques were studied, and their performance was evaluated in the context of the WPF problem. The aim was not to build a "complete" WPF model but to evaluate the forecast capability of those techniques. Generally, these techniques are used to convert the NWP forecasts to wind power —the so-called "wind-to-power (W2P)" model.

Fugon et al. [139] presented a survey on the performance of different data-mining models in WPF. Two versions of linear regression were examined: one is a simple regression model used as reference, and the other consists of combining the input variables to create extra variables. The analyzed nonlinear models were NN, SVM [140], regression trees with bagging, and random forests for regression [141]. The performance of each model was assessed in three wind farms located in France for a time horizon of 60 hr. All models outperformed persistence, and a global superiority of the nonlinear models was verified in the three wind farms. However, the performance of the linear models is reasonably good when compared with the persistence model. The nonlinear model with best results was the random forests model. The random forests are a combination of tree predictors, where each tree depends on the values of a random vector independently sampled with the same distribution for all trees in the forest. The excellent performance of this model means that, as found in other papers mentioned previously, using multiple models for WPF may decrease the forecast error.

Negnevitsky et al. [142] and [143] addressed the combined use of neural networks and fuzzy logic in WPF. This is a hybrid approach called Adaptive Neural Fuzzy System (ANFIS). 50 Although this model has only been applied for very short-term forecasting, and although the authors only present results for this case, it can be applied for time horizons of between 24 and 72 hr.

Jursa [144] compares different techniques for wind power forecasts, such as a classical MLP NN, mixture of experts [145], SVM, and nearest neighbor search with a Particle Swarm Optimization (PSO) algorithm for feature selection of the input of several locations in a spread area [146]. The author additionally combines different models by averaging the model outputs. The results for 10 wind farms located in Germany were compared, and NWPs were available for each wind farm. The best model was the ensemble with three different models (i.e., mixture of experts, nearest neighbor, and SVM), with a 15% improvement over an NN. Using all the models in the ensemble, however, is not always the best solution. In fact, the ensemble with all four models in some wind farms has registered a lower improvement when compared with the three-model ensemble. The best individual model was the mixture-of-experts model, which achieved an 8.8% improvement over the NN. The results of the SVM are always better when compared with the neural network results. The nearest-neighbor model was better than the NN in some wind farms. However, in others, the NN performed better. The results showed the advantages of combining several models for day-ahead forecasts. The intention with this research was to improve the commercial model Wind Power Management System (WPMS) developed by the Institut für Solare Energieversorgungstechnik (ISET).

Kusiak et al. [147] tested five data-mining models to produce forecasts for very shortterm horizons (1 to 12 hr ahead) and short-/long-term horizons (3 to 84 hr ahead) by using NWP forecasts from the Rapid Update Cycle (RUC) model and the North American Mesoscale (NAM) model, respectively. Two different forecast methodologies have bean compared and analyzed: (i) a direct forecast method, in which the power generation is forecasted directly from the weather forecasting data; and (ii) an integrated method, in which the wind speed is forecasted with the weather data, and then the power is forecasted with the predicted wind speed with a knearest neighbor algorithm. The authors used a boosting tree algorithm [148] for selecting the most relevant NWP data points surrounding the wind farm. Moreover, even after feature selection, a principal component analysis (PCA) analysis is applied to reduce the input dimensionality.

The five data-mining models were: SVM, MLP NN, RBF NN, regression trees, and random forests. The MLP NN outperforms the other four models in both very-short and short- /long-term forecasts. The direct approach outperformed the integrated approach also for both very-short and short-/long-term.

The authors stressed the strong dependence between the WPF model accuracy and the NWP model accuracy.

Jursa and Rohrig [149] presented an approach for one-hour-ahead forecasts (we believe that this approach can be applied to day-ahead forecasts without major modifications) based on the application of optimization algorithms for feature selection and models parameter optimization. In order to forecast for a single wind farm, the authors used measured wind power data of several other wind farms (30 wind farms), as well as the NWP data of the corresponding 51 forecast points closest to the location of the wind farms. The two main contributions of the paper are: (i) a method that uses the spatial and temporal information from a wide area in order to improve a single wind farm forecast; and (ii) use of PSO and differential evolution [150] for the automatic selection of the input variables and model parameters. The two used forecast models were an NN and a nearest-neighbor search. The authors concluded that the wind power forecast error can be reduced with the use of optimization algorithms for feature selection and parameter setting. With this approach, it is possible to reduce forecast error for most wind farms, in comparison to the manually set NN model. For example, the mean improvement in the forecast error in comparison to the persistence of the best model approach (i.e., NN automated specified using PSO) was 9.6%, while the percentage was 6.8% with the manually specified NN. In combining the NN and the nearest-neighbor approaches, there was a 10.75% improvement.

Duran et al. [130] studied the ARX models for very-short-term and short-term forecasting with NWP and on-line generation data as inputs. They noticed that the improvement obtained in short time horizons (i.e., of 6 hr) was smaller as a result of the increasing relevance of the past output power when compared with the NWP for this time horizon. One the other hand, when the time horizon is 24 hr ahead, the past output power loses importance, and the NWP gains more relevance. When compared with the AR, the improvement of the ARX is about 14% for 24 hr and about 26% when compared with the persistence.

Barbounis and Theocharis [151] and [152] employed locally recurrent neural networks to forecast wind speed and power 72 hr ahead, based on meteorological information. Three types of local recurrent neural networks were studied: (i) the infinite impulse response multilayer perceptron (IIR-MLP); (ii) the local activation feedback multilayer network (LAF-MLN); and (iii) and the diagonal recurrent neural network (RNN). Two new and efficient learning algorithms are presented: a global and a decoupled approach of the recursive prediction error (RPE) algorithm to train the neural network on-line (i.e., by updating weights and bias on-line). In the global RPE (GRPE), all weights are simultaneously updated. Moreover, to cope with the increased computational complexity of the GRPE, a local version of the algorithm, called a decoupled RPE (DRPE), was developed. The DRPE consists of dividing the global optimization problem into a set of manageable subproblems at the neuron level. In so doing, it is possible to reduce the computational and storage requirements considerably, while preserving high accuracy qualities of the GRPE at the same time. The three recurrent networks were compared to two static models, a finite-impulse response NN (FIR-NN) and a conventional static-MLP network. The performance of the proposed models was tested on a wind farm located in the Greek island of Crete, and the NWPs were provided by the regional forecast SKIRON for four points that were 30 km away from the wind farm. The results showed that the recurrent networks performed better when compared to the static models in all look-ahead times. The FIR-NN outperformed the static MLP by 11.82% and 12.7% in terms of mean absolute error. The recurrent models also achieved an average improvement of 50% (for look-ahead times longer than 20 hr) over persistence. Similar results are valid for the wind speed forecast. Because of the richness of the network architecture, IIR-MLP presented the best performance when compared with the other two. The main contributions of this paper are the new on-line training neural networks and algorithms.

These new training algorithms provide the ability to cope with changes in the wind farm behavior and operation, as well as low computational effort. Bessa et al. [153] reported the use of the back-propagation algorithm to directly train a neural network on-line. The methodology is as follows: (i) train a neural network by using a batch back-propagation approach with the available historical data; (ii) in the on-line mode, the NN makes predictions for time sample t+k at the time stamp t; and (iii) when the measured value is known, past k time then stamps the neural network forecasts again for time stamp t, and the forecast error that took place in t (on the new measured value) is computed and back-propagated through the network (weights and bias are updated) only once. This methodology makes it possible to deal adequately with data streams in the presence of concept drift or concept changes, which occur in the behavior of the wind. At the same time, these concept changes are also the result of the practical operations of wind farms, being caused by variations in the available generating capacity either because of maintenance or failure or simply because of capacity additions. The results from two wind farms located in Portugal have shown that there was an improvement with the use of on-line neural network training, not only in normal operation, but also in the event of a concept change.

It is well known that the wind speed vs. power curve of a wind turbine is highly nonlinear. The transformation of wind speed into wind power changes the statistical properties of the errors. This result has been shown, for instance, in [154] for six sites in Germany, where error distributions from wind power prediction models were skewed right and had a positive excess of kurtosis. This means that they were asymmetrical; they presented a higher frequency of errors to the left of the mean and were flatter than the Gaussian distribution. The authors stated the following: "The relative standard deviation of the measured power output is by a factor 1.8–2.6 larger than the relative standard deviation, of the time series of the wind speed measurement. This factor was caused by the power curve and can be regarded as the effective nonlinearity factor that describes the scaling of variations in the wind speed due to the local slope of the power curve… ." The same shape of error distribution can be found in [155],[156].

When observing the literature, it is possible to understand that, one way or another, models depend on a training process and usually adopt the Minimum Square Error (MSE) as a quality criterion. The applicability of MSE to train a mapper (any model mapping an inputoutput relation, such as neural networks, fuzzy inference systems, time series or other, with parameters to be learned) is only optimal if the probability distribution function (pdf) of the prediction errors is Gaussian [113]. Minimizing the square error is equivalent to minimizing the variance of the error distribution. Using this criterion, the higher moments (e.g., skewness, kurtosis) are not captured. However, they contain information that should be passed on to the parameters (weights) of the neural network instead of remaining in the error distribution.

The presence of non-Gaussian distributions has encouraged further research for new techniques to train mappers. In WPF, having a good mapper, means that it is possible to provide better estimates of the W2P model. To deal with this problem, alternative cost functions have been developed.

Pinson et al. [157] and [158] described two approaches for the local polynomial regression. The first work showed that the application of the recursive least squares in order to estimate the coefficients may lead to inaccurate results, especially in WPF where there is a significant amount of noise and a high number of outliers. An adaptive asymmetrical (yet convex) M-type estimator (inspired by the M-estimator described by Huber [159]) was proposed to deal with non-Gaussian errors. A local M-type estimator was proposed to account for the weighting in local polynomial regression. An innovative characteristic of this robust estimation is that, instead of defining a threshold value to reject outliers, a proportion of suspicious residuals is a model parameter. Another innovative contribution has to do with the time-varying coefficients, which allow the model to cope with the nonstationary process. The model was tested on datasets that included wind speed measurements at the level of a wind farm, as well as simulated wind speed and power data. The simulation results have shown that using the M-estimator and the local robustification is quite advantageous.

The second work describes an approach to compute the local linear coefficients that are orthogonally fitted by minimizing a weighted Total Least Squares (TLS). The main core of the approach has to do with the local minimization of the Euclidian distance between observations and the estimated nonparametric regression function. The robustification of the method follows the M-type estimation described above for local polynomial regression. Furthermore, an adaptive estimation of the coefficients was introduced to deal with the non-stationary process. This method was evaluated on semi-artificial data and the real wind speed data was passed through a modeled power curve in order to obtain noise free power data. Then, both wind speed and power data were corrupted in order to generate realistic datasets of wind speed forecasts and the corresponding power measurements. The comparison of the proposed method and the least square (LS) model showed a significant improvement. For instance, the value of the NMAE compared against noise-free data is 2.46% for the LS, 1.08% for the orthogonal fitting, and 0.97% for the robust orthogonal fitting. Moreover, a comparison of the final W2P conversion process obtained with the LS and robust TLS indicated that the true W2P conversion process became closer to the TLS method.

A different cost function consists of minimizing the information content of the error distribution instead of minimizing its variance (MSE). Entropy is a measure of information content and the incorporation of entropy as a cornerstone concept in mapper training has been the object of Information Theoretic Learning (ITL) [160],[161]. Miranda et al. [162] and Bessa et al. [163], in a first application devoted to wind power forecast, used an Evolutionary Particle Swarm Optimization (EPSO) algorithm to carry out an off-line optimization of the weights of a Takagi-Sugeno FIS, which is worked as a W2P model. In that paper, a comparison was established between a Takagi-Sugeno FIS trained by MSE and one trained by minimizing Renyi's quadratic entropy [161] of the error distribution — and the results have shown that a higher frequency of errors close to zero was produced by the entropybased model. In another paper, Bessa et al. [164] engaged in evaluating the performance of neural networks that were trained in off-line mode by comparing the MSE criterion with three 54 ITL-inspired criteria. The conclusion that was drawn from the analysis of two real cases of Portuguese wind farms, a 21.6 MW farm (12 units of 1.8 MW each) and a 16.2 MW farm (17 units of 0.6 MW and 3 units of 2 MW), was unmistakable: in off-line training, entropy, as a performance criterion, leads to better predictions (in terms of frequent errors close to zero and insensitivity to outliers), as opposed to the adoption of MSE as a training criterion. The error distribution of the MSE and two ITL-inspired criteria, Minimim Error Entropy (MEE) and Maximum Correntropy Criteria (MCC), were tested on the two Portuguese wind farms. It was possible to obtain a narrower pdf (probability distribution function) with MEE and MCC criteria than with MSE. In fact, if the error pdfs were Gaussian, the MSE criterion would perform as well as an entropy-based criterion, but that was not the case. Therefore, according to the theory, it was possible to design a mapper that would produce a predictor with a higher frequency of errors close to zero, a characteristic that is associated with the smaller entropy of the pdf. Generally, this is desirable. Bessa et al. [153] presented practical results supporting two ideas: (i) criteria based on entropy (i.e., the measure of information content) of the distribution of prediction error are more suitable than the traditional MSE criterion in order to train accurate wind power prediction models; and (ii) the entropy-based criteria can be formatted into on-line adaptive models that perform better than off-line trained models when using feed-forward NN. The test was performed for two wind farms located in Portugal, and the results have shown that: (i) the NMAE of the ITL criteria is below the one obtained with the MSE criterion; and (ii) the advantages of using an on-line training are higher. Furthermore, in uncertainty forecasting, the non-Gaussian shape of error distribution was also addressed. More details regarding the modeling and estimation of uncertainty will be provided in Chapter 5.

Salcedo-Sanz et al. [165] presented a hybridization of the MM5 model with a neural network that tackles the final statistical downscaling process to obtain the wind speed forecast for each wind turbine of a wind farm. The MM5 model performs a physical downscaling of the data from the global model to obtain a wind speed forecast for a small area grid. In the second downscaling process (statistical), an NN takes as inputs the following variables: wind speed forecasted of two grid points surrounding the wind farm, wind direction in one of the points, temperature in one of the points, and two solar cycle equations. The output of the NN is the forecasted wind speed for each wind turbine.