9

Modeling Financial Wind Derivatives

Weather derivatives are financial tools that can help organizations or individuals to reduce risk associated with adverse or unexpected weather conditions and can be used as part of a risk management strategy. Weather derivatives linked to various weather indices, such as rainfall, temperature, or wind, are traded extensively in CME as well as on the OTC market. The electricity sector is especially sensitive to the temperature and wind since temperature affects the consumption of electricity while wind affects production of electricity in wind farms. Hence, it is logical that energy companies are the main investors of the weather market. In Chapter 8 a detailed framework for modeling and pricing temperature derivatives was developed. In this chapter we focus on wind derivatives.

The notional value of the wind-linked securities traded is around $36 million, indicating a large and growing market (WRMA, 2010). However, after the close of the U.S. Future Exchange, wind derivatives have been traded OTC. A demand for these derivatives exists. However, investors hesitate to enter into wind contracts. The main reasons for the slow growth of the wind market compared to temperature contracts are the difficulty in modeling wind accurately and the challenge of finding a reliable model for valuing related contracts. As a result, there is a lack of reliable valuation framework that makes financial institutions reluctant to quote prices over these derivatives.

The aim of this chapter is to model and price wind derivatives. Wind derivatives are standardized products that depend only on the daily average wind speed measured

by a predefined meteorological station over a specified period and can be used by wind (and weather in general)-sensitive businesses such as wind farms, transportation companies, construction companies, and theme parks. The financial contracts that are traded are based on the simple daily average wind speed index, and this is the reason that we choose to model only the dynamics of the daily average wind speeds. The revenues of each company have a unique dependence and sensitivity to wind speeds. Although wind derivatives and weather derivatives can hedge a significant part of the weather risk of the company, some basis risk will always still exist which must be hedged from each company separately. This can be done either by defining a more complex wind index or by taking an additional hedging position. Cao and Wei (2003) provided various examples of how weather derivatives can reduce the basis and volumetric risks in weather-sensitive businesses (e.g., ski resorts, restaurants, theme parks, electricity companies).

Wind is a free, renewable, and environmentally friendly source of energy (Billinton et al., 1996). Wind derivatives are traded extensively in the electricity sector. While the demand for electricity is closely related to the temperature, the electricity produced by a wind farm is dependent on the wind conditions. The risk exposure of the wind farm depends on the wind speed, the wind direction, and in some cases the wind duration of the wind speed at certain levels. However, modern wind turbines include mechanisms that allow turbines to rotate in the appropriate wind direction (Caporin and Pres, 2010). Wind derivatives can be used as part of a hedging-strategy in wind-sensitive businesses. However, the underlying wind indices do not account for the duration of the wind speed at certain levels but, rather, usually measure the average daily wind speed. Hence, the parameter of the duration of the wind speed at certain levels is not considered in our daily model. Hence, the risk exposure of a wind-sensitive company can be analyzed by quantifying only the wind speed. On the other hand, companies such as wind farms whose revenues depend on the duration effect can use an additional hedging strategy that includes this parameter. This can be done by introducing a second index that measures the duration. A similar index for temperature is the frost day index.

Many different approaches have been proposed so far for modeling the dynamics of the wind speed process. The most common is the generalized autoregressive moving average (ARMA) approach. There has been a number of studies on the use of linear ARMA models to simulate and forecast wind speed in various locations (Billinton et al., 1996; Caporin and Pres, 2010; Castino et al., 1998; Daniel and Chen, 1991; Huang and Chalabi, 1995; Kamal and Jafri, 1997; Martin et al., 1999; Saltyte-Benth and Benth, 2010; Tol, 1997; Torres et al., 2005). Kavasseri and Seetharaman (2009) used a more sophisticated fractional integrated ARMA (ARFIMA) model. Most of these studies did not consider in detail the accuracy of the wind speed forecasts (Huang and Chalabi, 1995). On the other hand, Ailliot et al. (2006) apply an autoregressive model (AR) with time-varying coefficients to describe the space-time evolution of wind fields. Benth and Saltyte-Benth (2009) introduced a stochastic process called the continuous AR (CAR) model to model and forecast daily wind speeds. Finally, Nielsen et al. (2006) presented various statistical methods for short-term wind speed forecasting. Sfetsos (2002) argues about the use of linear or meteorological models, since their prediction error is not significantly lower than

the elementary persistent method. Alternatively, some studies use space-state models to fit the speed and the direction of the wind simultaneously (Castino et al., 1998; Cripps et al., 2005; Haslett and Raftery, 1989; Martin et al., 1999; Tolman and Booij, 1998; Tuller and Brett, 1984).

Alternative to the linear models, artificial intelligence was used for wind speed modeling and forecasting. Alexiadis et al. (1998), Barbounis et al. (2006), Beyer et al. (1994), Mohandes et al. (1998), More and Deo 2003, and Sfetsos (2000, 2002) utilized neural networks to model the dynamics of the wind speed process. Mohandes et al. (2004) used support vector machines, and Pinson and Kariniotakis (2003) employed fuzzy neural networks.

Depending on the application, wind modeling is based on an hourly (Ailliot et al., 2006; Castino et al., 1998; Daniel and Chen, 1991; Kamal and Jafri, 1997; Martin et al., 1999; Sfetsos, 2000, 2002; Torres et al., 2005; Yamada, 2008), daily (Benth and Saltyte-Benth, 2009; Billinton et al., 1996; Caporin and Pres, 2010; Huang and Chalabi, 1995; More and Deo, 2003; Tol, 1997), weekly, or monthly basis (More and Deo, 2003). When the objective is to hedge against electricity demand and production, hourly modeling is used, whereas for weather derivative pricing the daily method is used. More rarely, weekly or monthly modeling is used to estimate monthly wind indexes. Since we want to focus on weather derivative pricing, the daily modeling approach is followed; however, the method proposed can be easily be adapted to hourly modeling.

Wind speed modeling is much more complicated than temperature modeling since wind has a direction and is greatly affected by the surrounding terrain, such as buildings, and trees (Jewson et al., 2005). However, Benth and Saltyte-Benth (2009) have shown that wind speed dynamics share a lot of common characteristics with the dynamics of temperature derivatives. In this context we use a mean reverting Ornstein–Uhlenbeck stochastic process to model the dynamics of the wind speed, where the innovations are driven by a Brownian motion. The statistical analysis reveals seasonality in the mean and variance. In addition we use a novel approach to model the autocorrelation of wind speeds. More precisely, a wavelet network is utilized to capture accurately the autoregressive characteristics of the wind speeds.

The evaluation of the proposed methodology against alternative modeling procedures proposed in prior studies indicates that wavelet networks can accurately model and forecast the dynamics and evolution of the speed of the wind. The performance of each method was evaluated in-sample as well as out-of-sample and for different time periods.

The remainder of the chapter is organized as follows. First, a statistical analysis of the wind speed dynamics is presented. Then, a linear model and a nonlinear nonparametric wavelet network are fitted to the data. Next, an evaluation and a comparison of the models studied is presented. Finally, we conclude.

MODELING THE DAILY AVERAGE WIND SPEED

In this section we derive empirically the characteristics of the daily average wind speed (DAWS) dynamics in New York. The data were collected from NOAA
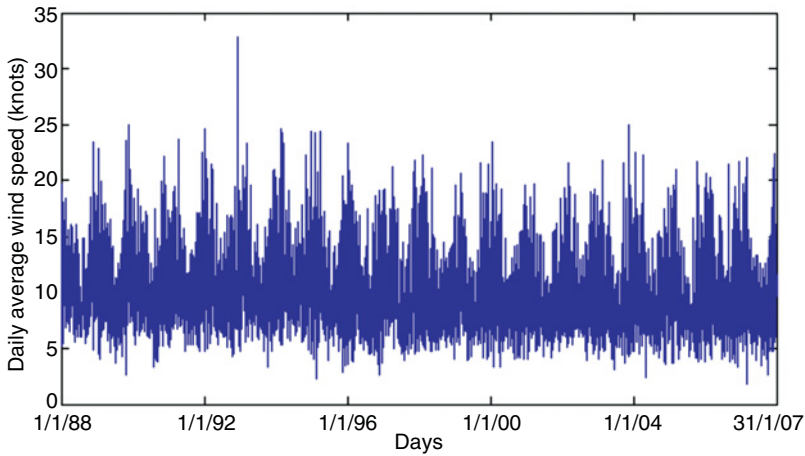
Figure 9.1   *Daily average wind speed for New York.*

(http://www.noaa.gov/) and correspond to DAWSs. The wind speeds are measured in units of 0.1 knot. The measurement period is between January 1, 1988 and February 28, 2008. The first 20 years are used for estimation of the parameters, while the remaining two months are used for evaluation of the performance of the model proposed. So that each year will have the same number of observations, February 29 is removed from the data, resulting in 7359 data points. The time series is complete without any missing values.

In Figure 9.1 the DAWSs for the first 20 years are presented. The descriptive statistics of the in-sample data are presented in Table 9.1. The values of the data are always positive and range from 1.8 to 32.8 with a mean around 9.91. Also, a closer inspection of Figure 9.1 reveals seasonality.

The descriptive statistics of the DAWSs indicate that there is a strong positive kurtosis and skewness, while the normality hypothesis is rejected based on the Jarque–Bera statistic. The same conclusions can be reached by observing the first part of Figure 9.2, where the histogram of the DAWSs is represented. The distribution of DAWSs deviates significantly from the normal and is not symmetrical. In the literature the Weibull or the Rayleigh (which is a special case of the Weibull) distributions were proposed describe the distribution of the wind speed (Brown et al., 1984; Daniel and Chen, 1991; Garcia et al., 1998; Justus et al., 1978; Kavak Akpinar and Akpinar, 2005; Nfaoui et al., 1996; Torres et al., 2005; Tuller and Brett, 1984). In addition, some

TABLE 9.1   Descriptive Statistics of the Wind Speed in New York[a]

	Mean	Med.	Max.	Min.	S. Dev.	Skew	Kurt.	JB	*p*-Value
Original	9.91	9.3	32.8	1.8	3.38	0.96	4.24	1595.41	0
Transformed	2.28	2.3	3.6	0.6	0.34	0.00	3.04	0.51	1

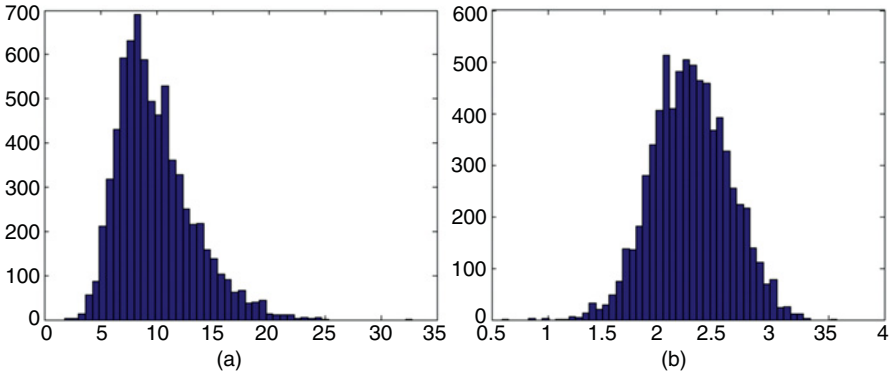[a]JB, Jarque–Bera statistic; *p*-value, *p*-values of the JB statistic.

Figure 9.2 *Histogram of the (a) original and (b) Box–Cox transformed data.*

studies propose use of the lognormal distribution (Garcia et al., 1998) or chi-square (Dorvlo, 2002). Finally, Jaramillo and Borja (2004) used a bimodal Weibull and Weibull distribution. However, empirical studies favor use of the Weibull distribution (Celik 2004; Tuller and Brett, 1984).

A closer inspection of Figure 9.2a reveals that the DAWSs in New York follow a Weibull distribution with scale parameter $\lambda = 11.07$ and shape parameter $k = 3.04$. To symmetrize the data, the Box–Cox transform is applied. The Box–Cox transformation is given by

$$W^{(l)} = \begin{cases} \frac{W^l - 1}{l} & l \neq 0 \\ \ln(W) & l = 0 \end{cases} \tag{9.1}$$

where $W^{(l)}$ is the transformed data. The parameter l is estimated by maximizing the log-likelihood function. Note that log-transform is a special case of the Box–Cox transform with $l = 0$. The optimal l of the Box–Cox transform for the DAWSs in New York is estimated to be 0.014. In Figure 9.2b the histogram of the transformed data can be found while the second row of Table 9.1 shows the descriptive statistics of the transformed data.

The DAWSs exhibit a clear seasonal pattern which is preserved in the transformed data. The same conclusion can be reached by examining the autocorrelation function (ACF) of the DAWSs in the first part of Figure 9.3. The seasonal effects are modeled by a truncated Fourier series, given by

$$S(t) = a_0 + b_0 t + \sum_{i=1}^{I_1} a_i \sin \frac{2\pi i(t - f_i)}{365} + \sum_{j=1}^{J_1} b_j \sin \frac{2\pi j(t - g_j)}{365} \tag{9.2}$$

In addition, we examine the data for a linear trend representing the global warming or the urbanization around the meteorological station. First, we quantify the trend by fitting a linear regression to the DAWS data. The regression is statistically significant
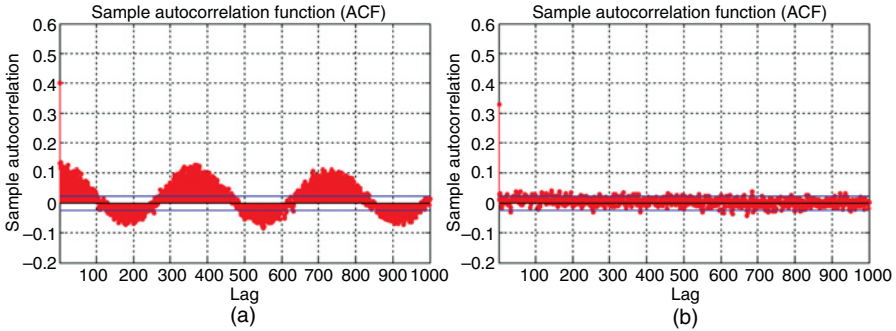
Figure 9.3 *Autocorrelation function of the transformed DAWSs in New York (a) before and (b) after removing the seasonal mean.*

TABLE 9.2   Estimated Parameters of the Seasonal Component

a_0	b_0	a_1	f_1	b_1	g_1
2.3632	−0.000024	0.0144	827.81	0.1537	28.9350

with intercept $a_0 = 2.3632$ and slope $b_0 = -0.000024$, indicating a slight decrease in the DAWSs. Next, the seasonal periodicities are removed from the detrended data. The remaining statistically significant estimated parameters of equation (9.2) with $I_1 = J_1 = 1$ are presented in Table 9.2. As shown in Figure 9.3b, the seasonal mean was removed successfully. The same conclusion was reached in previous studies for daily models for both temperature and wind (Benth and Saltyte-Benth, 2009; Zapranis and Alexandridis, 2008).

LINEAR ARMA MODEL

In the literature, various methods have been proposed for studying the statistical characteristics of the wind speed, in daily or hourly measurements. However, the majority of the studies utilize variations of the general ARMA model. In this chapter we first estimate the dynamics of the detrended and deseasonalized DAWS process using a general ARMA model and then compare our results with a wavelet network.

We define the detrended and deseasonalized DAWS as

$$\tilde{W}^{(l)}(t) = W^{(l)}(t) - S(t) \tag{9.3}$$

The dynamics of $\tilde{W}^{(l)}(t)$ are modeled by a vectorial Ornstein–Uhlenbeck stochastic process,

$$d\tilde{W}^{(l)}(t) = \mathbf{a}\tilde{W}^{(l)}(t)\,dt - I_p\sigma(t)\,dB_t \tag{9.4}$$

where I_p is the pth unit vector in R^p and $\sigma(t)$ is the standard deviation.
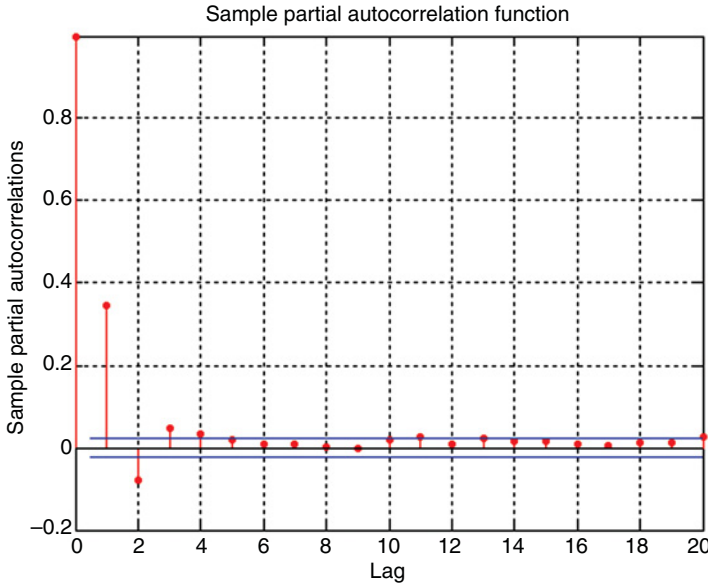
Sample partial autocorrelation function

Figure 9.4   *Partial autocorrelation function of the detrended and deseasonalized DAWSs in New York.*

First, to select the correct ARMA model, we examine the ACF of the detrended and deseasonalized DAWSs. A closer inspection of Figure 9.3b reveals that the lags 1, 2, and 4 are significant. On the other hand, by examining the partial autocorrelation function (PACF) in Figure 9.4, we conclude that the first four lags are necessary to model the autoregressive effects of the dynamics of the wind speed.

To find the correct model, we estimate the log-likelihood function (LLF) and the akaike information criterion (AIC). Consistent with the PACF, both criteria suggest that an AR(4) model is adequate for modeling the wind process since they were minimized when a model with four lags was used. The estimated parameters and the corresponding *p*-values are presented in Table 9.3. It is clear that the three first parameters are statistically very significant since their *p*-value is less than 0.05. The parameter of the fourth lag is statistically significant with a *p*-value of 0.0657. The AIC for this model is 0.46852 and the LLF is −1705.14.

Observing the residuals of the AR model in Figure 9.5a, we conclude that the autocorrelation was removed successfully. However, the ACF of the squared residuals indicates a strong seasonal effect in the variance of the wind speed, as shown in Figure 9.6. Similar behavior was observed in the residuals of temperature and wind in

TABLE 9.3   Estimated Parameters of the Linear AR(4) Model

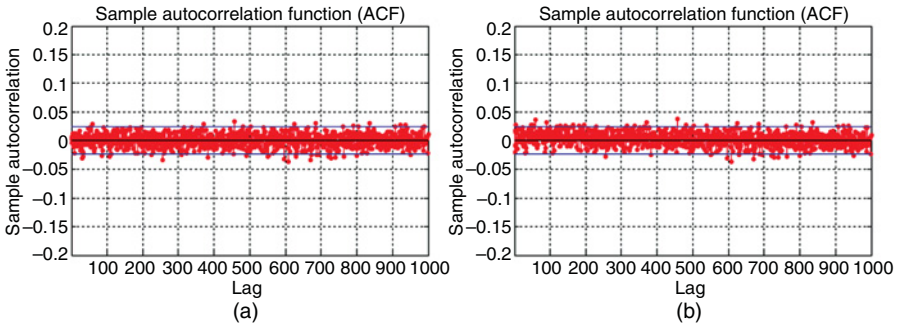| Parameter	AR(1)	AR(2)	AR(3)	AR(4)
Value	0.3617	−0.0999	0.0274	0.0216
p-Value	0.0000	0.0000	0.0279	0.0657

Figure 9.5 *Autocorrelation function of the residuals of (a) the linear model and (b) the WN.*

various studies (Zapranis and Alexandridis, 2008). The seasonal variance is modeled with a truncated Fourier series:

$$\sigma^2(t) = c_0 + \sum_{i=1}^{I_2} c_i \sin \frac{2i\pi t}{365} + \sum_{j=1}^{J_2} d_j \sin \frac{2\pi jt}{365} \tag{9.5}$$

Note that we assume that the seasonal variance is periodic and repeated every year [i.e., $\sigma^2(t + 365) = \sigma^2(t)$, where $t = 1, \ldots, 7359$]. The empirical and fitted seasonal variance are presented in Figure 9.7 and the estimated parameters of equation (9.5) are presented in Table 9.4.

Not surprisingly, the variance exhibits the same characteristics as in the case of temperature. More precisely, the seasonal variance is higher in the winter and early summer, while it reaches its lower values during the summer period.

Finally, the descriptive statistics of the final residuals are examined. A closer inspection of Table 9.5 shows that the autocorrelation has been removed successfully, as indicated by the Ljung–Box Q-statistic. In addition, the distribution of the residuals is very close to the normal distribution, as shown in Figure 9.8a. However, small
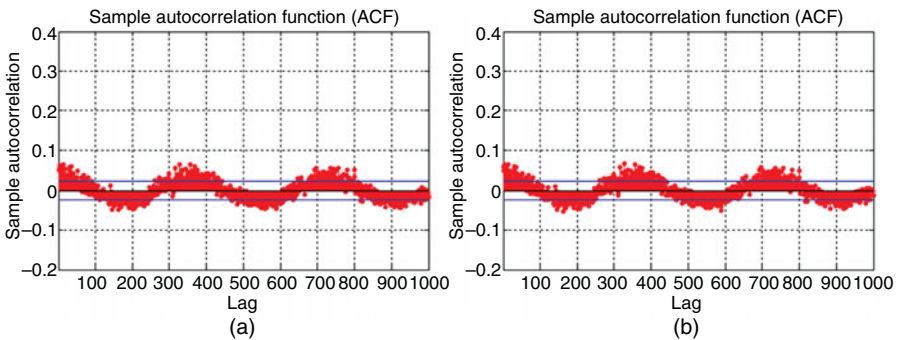

Figure 9.6 *Autocorrelation function of the squared residuals of (a) the linear model and (b) the WN.*
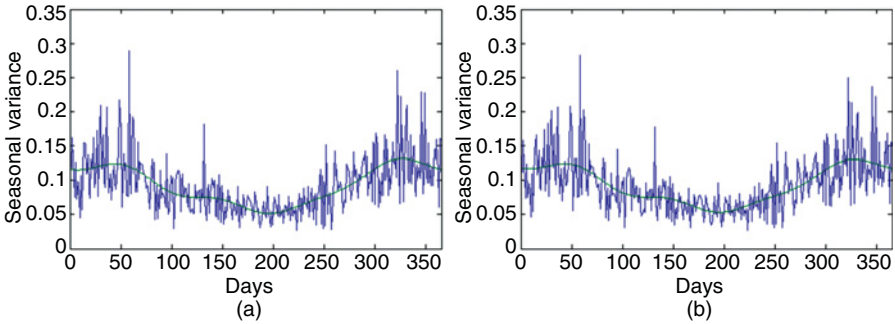
Figure 9.7   *Empirical and fitted seasonal variance of (a) the linear model and (b) the WN.*

TABLE 9.4   Estimated Parameters of the Seasonal Variance in the Linear Model

c_0	c_1	c_2	c_3	c_4	d_1	d_2	d_3	d_4
0.0932	0.000032	−0.0041	0.0015	−0.0028	0.0358	−0.0025	−0.0048	−0.0054

TABLE 9.5   Descriptive Statistics of the Residuals for the Linear AR(4) Model

Var.	Mean	S. dev.	Max.	Min.	Skew	Kur.	JB	p-Value	KS	p-Value	LBQ	p-Value
Noise	0	1	3.32	−5.03	−0.09	3.03	10.097	0.007	1.033	0.2349	8.383	0.989

S. dev., standard deviation; JB, Jarque–Bera statistic; KS, Komogorov–Smirnov statistic; LBQ, Ljung–Box Q–statistic.

negative skewness exists. More precisely, the residuals have mean 0 and standard deviation 1. In addition, the kurtosis is 3.03 and the skewness is −0.09.

Concluding, the previous analysis indicates that an AR(4) model provides a good fit for the wind process, while the final residuals are very close to the normal distribution.
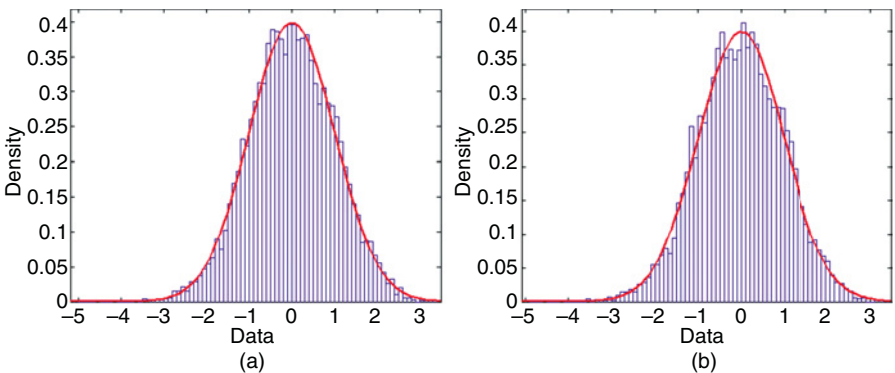

Figure 9.8   *Empirical and fitted normal distribution of the final residuals of (a) the linear model and (b) the WN.*

WAVELET NETWORKS FOR WIND SPEED MODELING

In this section, wavelet networks are applied in the transformed, detrended, and deseasonalized wind speed data in order to model the daily dynamics of wind speeds in New York. Motivated by the waveform of the data, we expect a wavelet function to better fit the wind speed. In addition, it is expected that the nonlinear form of the wavelet network will provide more accurate representation of the dynamics of the wind speed process both in-sample and out-of-sample.

In the context of the linear ARMA model, the mean reversion parameter a is typically assumed to be constant over time. Brody et al. (2002) mentioned that, in general, a should be a function of time, but no evidence was presented. The impact of a false specification of a on the accuracy of the pricing of temperature derivatives is significant (Alaton et al., 2002).

In this section we address that issue by using a wavelet network to estimate nonparametrically the relationship (9.4) and then estimate **a** as a function of time. In addition, we propose the variable selection algorithm presented in previous chapters for selecting the statistical significant lags. Hence, a series of speed of mean reversion parameters, $a_i(t)$, are estimated. By computing the derivative of the network output with respect to the network input, we obtain a series of daily values for $a_i(t)$. This is done for the first time, and it gives us a much better insight into DAWS dynamics and wind derivative pricing. As we will see, the daily variation of $a_i(t)$ is quite significant after all. In addition, it is expected that the waveform of the wavelet network will provide a better fit to the DATs that are governed by seasonalities and periodicities.

Using wavelet networks the generalized version of the autoregressive dynamics of detrended and deseasonalized DAWS estimated nonlinearly and nonparametrically is given by

$$\tilde{W}^{(l)}(t+1) = \phi\left(\tilde{W}^{(l)}(t), \tilde{W}^{(l)}(t-1), \ldots\right) + e(t) \tag{9.6}$$

where

$$e(t) = \sigma(t)\varepsilon(t) \tag{9.7}$$

and $\tilde{W}^{(l)}(t)$ is given by (9.3).

Variable Selection

Model (9.6) uses past DAWSs (detrended and deseasonalized) over one period. Using more lags, we expect to overcome the strong correlation found in the residuals. However, the length of the lag series must be selected. In previous chapters detailed explanations were given of how to apply the model identification framework. Model identification can be separated into two parts: model selection and variable significance testing. Since wavelet networks are nonlinear tools, criteria such as AIC or LLF cannot be used. Hence, in this section the framework proposed will be used to select the significant lags, to select the appropriate network structure, to train a wavelet network in order to learn the dynamics of the wind speeds, and finally, to forecast the future evolution of wind speeds.

The model identification algorithm uses a recurrent algorithm to simultaneously estimate the correct number of lags that must be used to model the wind speed dynamics and the architecture of the wavelet network.

Our backward selection algorithm examines the contribution of each available explanatory variable to the predictive power of the wavelet network. First, the prediction risk of the wavelet network is estimated as well as the statistical significance of each variable. If a variable is statistically insignificant, it is removed from the training set and the prediction risk and new statistical measures are estimated. The algorithm stops if all explanatory variables are significant. Hence, in each step of our algorithm, the variable with the larger p-value greater than 0.1 will be removed from the training set of our model. After each variable removal, a new architecture of the wavelet network will be selected and a new wavelet network will be trained. However, the correctness of the decision of removing a variable must be examined. This can be done by examining either the prediction risk or $\bar{R}^2$. If the new prediction risk is smaller than the new prediction risk multiplied by a threshold, the decision of removing the variable was correct. If the prediction risk increased more than the allowed threshold, the variable was reintroduced back to the model. We set this threshold at 5%. The statistical measure selected is the SBP. Our results indicate that the SBP fitness criterion was found to outperform alternative criteria significantly in the variable selection algorithm.

The variable selection framework proposed will be utilized for the transformed, detrended, and deseasonalized wind speeds in New York to select the length of the lag series. The target values of the wavelet network are the DAWSs. The explanatory variables are lagged versions of the target variable. The relevance of a variable to the model is quantified by the SBP criterion. Initially, the training set contains the dependent variable and seven lags. The analysis in the preceding section indicates that a training set with seven lags will provide all the necessary information of the ACF of the detrended and deseasonalized DAWSs. Hence, the initial training set consists of 7 inputs, 1 output, and 7293 training pairs.

Table 9.6 summarizes the results of the model identification algorithm for New York. Both the model selection and the variable selection algorithms are included in the table. The algorithm concluded in four steps and the final model contains only three variables (i.e., three lags). The prediction risk for the reduced model is 0.0937,

TABLE 9.6   Variable Selection with Backward Elimination in New York[a]

Step	Variable to Remove (Lag)	Variable to Enter (Lag)	Variables in Model	Hidden Units (Parameters)	n/p Ratio	Empirical Loss	Prediction Risk
—			7	1 (23)	317.4	0.0467	0.0938
1	7	—	6	1 (20)	365.0	0.0467	0.0940
2	5	—	5	1 (17)	429.4	0.0467	0.0932
3	6	—	4	2 (23)	317.4	0.0467	0.0938
4	4	—	3	2 (18)	405.6	0.0468	0.0937

[a]The algorithm concluded in four steps. In each step the following are presented: which variable is removed, the number of hidden units for the particular set of input variables and the parameters used in the wavelet network, the ratio between the parameters and the training patterns, the empirical loss, and the prediction risk.

while for the original model it was 0.0938, indicating that the predictive power of the wavelet network was slightly increased. On the other hand, the empirical loss increased slightly from 0.0467 for the initial model to 0.0468 for the reduced model, indicating that the explained variability (unadjusted) decreased slightly. Finally, the complexity of the network structure and the number of parameters were reduced significantly in the final model. The initial model needed 1 hidden unit and 7 inputs. Hence, 23 parameters were adjusted during the training phase, so the ratio of the number of training pairs n to the number of parameters p was 317.4. In the final model only 2 hidden units and 3 inputs were used, so only 18 parameters were adjusted during the training phase and the ratio of the number of training pairs n to the number of parameters p was 405.6.

The statistics for the wavelet network at each step are given in Table 9.7. The first part of the Table 9.7 reports the values of the SBP and its p-value; then various fitting criteria are reported. A closer inspection of the table reveals that the various error measures are reduced in the final model. However, the values of $\bar{R}^2$ are relatively small in all cases. This is due to the presence of large noise values compared to the small values of the underlying function.

In the final model, only three of the seven variables were used. The complexity of the model was reduced while the prediction power of the reduced model was increased. However, the in-sample obtained was a slightly poorer fit. The algorithm proposed suggests that a wavelet network needs only three lags to extract the autocorrelation from the data, whereas the linear model needed four lags. A closer inspection of Table 9.6 reveals that wavelet networks with three and four lags have the same predictive power in-sample and out-of-sample. Hence, we chose the simpler model.

TABLE 9.7   Step-by-Step Variable Selection in New York[a]

	Full Model		Step 1		Step 2		Step 3		Step 4	
Variable	SBP	*p*-Value	SBP	*p*-Value	SBP	*p*-Value	SBP	*p*-Value	SBP	*p*-Value
7	0.0000	**0.8392**								
6	0.0000	0.7467	0.0000	0.4855	0.0000	**0.9167**				
5	0.0000	0.6799	0.0000	**0.9467**						
4	0.0000	0.5203	0.0000	0.7180	0.0000	0.2643	0.0000	**0.7480**		
3	0.0001	0.1470	0.0001	0.0000	0.0001	0.4706	0.0001	0.4719	0.0003	0.0000
2	0.0010	0.0469	0.0010	0.0000	0.0010	0.0000	0.0009	0.0000	0.0010	0.0168
1	0.0141	0.0000	0.0141	0.0000	0.0137	0.0000	0.0140	0.0000	0.0135	0.0000
MAE	0.2430		0.2430		0.2428		0.2430		0.2429	
MaxAE	1.7451		1.7453		1.7156		1.7541		1.6986	
NMSE	0.8832		0.8832		0.8833		0.8832		0.8834	
$\bar{R}^2$	11.68%		11.67%		11.67%		11.68%		11.65%	
Empirical loss	0.0467		0.0467		0.0467		0.0467		0.0468	
Prediction risk	0.0938		0.0940		0.0932		0.0938		0.0937	
Iterations	22		37		26		19		225	

[a]The SBP is the average for each variable of 50 bootstrapped samples, the standard deviation, and the *p*-value; SBP, sensitivity-based pruning; MAE, mean absolute error; MaxAE, maximum absolute error; NMSE, normalized mean squared error; MSE, mean square error; MAPE, mean absolute percentage error.

TABLE 9.8   Prediction Risk at Each Step of the Variable Selection Algorithm for the First 5 Hidden Units for New York

Step	Hidden Units				
	1	2	3	4	5
0	**0.09378**	0.09380	0.09379	0.09379	0.09380
1	**0.09403**	0.09404	0.09403	0.09406	0.09406
2	**0.09321**	0.09324	0.09325	0.09326	0.09327
3	0.09384	**0.09380**	0.09384	0.09387	0.09386
4	0.09370	**0.09367**	0.09368	0.09373	0.09379

Model Selection

In this section the appropriate number of hidden units is determined by applying the model selection algorithm. Table 9.8 shows the prediction risk for the first 5 hidden units at each step of the variable selection algorithm for the DAWSs in New York. It is clear that only 1 hidden unit is sufficient to model the detrended and deseasonalized DAWSs in New York at the first three steps. Similarly, 2 hidden units were needed for the last two steps.

Initialization and Training

After the training set and the correct topology of the wavelet network are selected, the wavelet network can be constructed and trained. In this case study the BE method is used to initialize the wavelet network. A wavelet basis is constructed by scanning the first four levels of the wavelet decomposition of the detrended and deseasonalized DAWSs in New York.

The wavelet basis consists of 205 wavelets. To reduce the number of wavelets in the wavelet basis, the wavelets that contain fewer than six sample points of the training data in their support are removed. The truncated basis contains 119 wavelet candidates. Applying the BE method, the wavelets are ranked in order of significance. Since only 2 hidden units are used in the architecture of the model, the best two wavelets are selected. The results of previous steps are similar. The MSE after the initialization was only 0.09420. Figure 9.9a presented the initialization of the final model using 2 hidden units. The initialization is very good, and the wavelet network converged after 225 iterations. The training stopped when the minimum velocity, 10^{-5}, of the training algorithm was reached. The fitting of the trained wavelet network is shown in Figure 9.9b.

Model Adequacy

In this section the model adequacy of the wavelet network is studied. The n/p ratio is 405.3, indicating that each parameter of the network corresponds to 405 values. Hence, we can safely conclude that overfitting was avoided.
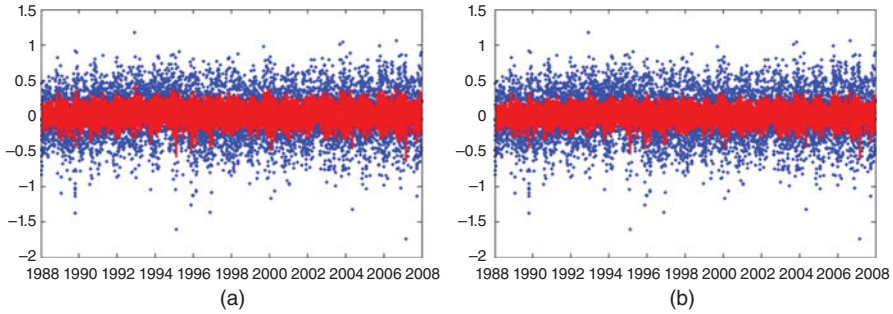
Figure 9.9   *Initialization of the final model for the wind data in New York using the BE method (a) and the fit of the trained network with 2 hidden units (b). The wavelet network converged after 225 iterations.*

TABLE 9.9   Residual Testing[a]

	Parameter	p-Values
n/p Ratio	405.3	–
Mean	0.0000	
Median	0.0018	
S. dev.	0.3058	
DW	2.0184	0.4526
LB *Q*-stat.	20.6002	0.4210
JB stat.	28.8917	0.0000
KS stat.	22.3453	0.0000

[a]S. dev., standard deviation; DW, Durbin–Watson; LB, Ljung–Box; KS, Kolmogorov–Smirnov.

In a closer examination of the residuals we found that the mean of the residuals is zero with a standard deviation of 0.3058. The normality hypothesis is rejected, but the hypothesis that the residuals are uncorrelated is accepted. The results are reported analytically in Table 9.9.

Various error criteria are reported in Table 9.10. The MSE is only 0.0937, while the NMSE and the RMSE are 0.8834 and 0.0937, respectively. Similarly, the maximum absolute error is 1.6986. Finally, the SMAPE is 27.55%. Note that since some values are zero, the MAPE cannot be computed. We observer that the SMAPE is relative high. This is due to the presence of a large error term compared to the very small values of the targets.

The parameters estimated for the regression between the target values and the network output are presented in Table 9.11. A closer inspection reveals that the

TABLE 9.10   Error Criteria[a]

Md.AE	MAE	MaxAE	SSE	RMSE	NMSE	MSE	MAPE	SMAPE
0.2034	0.2429	1.6986	682.4678	0.3058	0.8834	0.0937	—	27.55%

[a]Md.AE, median absolute error; MAE, mean absolute error; MaxAE, maximum absolute error; SSE, sum of squared errors; RMSE, root mean squared error; NMSE, normalized mean squared error; MSE, mean squared error; MAPE, mean absolute percentage error; SPAME, symmetric mean absolute percentage error.

TABLE 9.11   Regression Statistics[a]

	Parameter	p-Values	S.E.	T-Stat.
b_0	0.0001	0.9886	0.0036	0.0143
b_1	0.9655	0.0000	0.0311	31.0318
$b_1 = 1$ Test	0.9655	0.2670	0.0311	31.0318
F	962.9701	0.0000		
DW	1.9933	0.7611		

[a]S.E., squared error; DW, Durbin–Watson.

parameter b_0 is not statistically different from zero, while the parameter b_1 is not statistically different from 1 at significance level 5%. Moreover, the linear regression is statistically significant according to the F-statistic. Finally, as presented in Table 9.12, the changes in the direction metric are reported. More precisely, POCID, IPOCID, and POS are 69.64%, 43.28%, and 59.55%, respectively.

Speed of Mean Reversion and Seasonal Variance

The daily values of the speed of mean reversion function $a_i(t)$ (7293 values) are depicted in Figure 9.10. Since there are three significant lags, there are three mean-reverting functions. Our results indicate that the speed of mean reversion is not constant. On the contrary, its daily variation is quite significant; this fact naturally has an impact on the accuracy of the pricing equations and thus must be taken into account. Intuitively, it was expected that $a_i(t)$ would not be constant. If the wind speed today is away from the seasonal average, it is expected that the speed of mean reversion will be high (i.e., the wind speed cannot deviate from its seasonal mean for long periods).

Examining the second part of Figure 9.5, we conclude that the autocorrelation was removed from the data successfully; however, the seasonal autocorrelation in the squared residuals is still present, as shown in Figure 9.6. We remove the seasonal autocorrelation using equation (9.5). The estimated parameters are presented in Table 9.13 and, as expected, their values are similar to those of the case of the linear model. The empirical and fitted seasonal variance are presented in Figure 9.7. The variance is higher during the winter period and reaches its minimum during the summer period.

Finally, examining the final residuals of the wavelet network model, we observe that the distribution of the residuals is very close to the normal distribution shown in Figure 9.8 and the autocorrelation was removed from the data successfully. In addition, we observe an improvement in the distributional statistics, in contrast to the

TABLE 9.12   Change-in-Direction Metrics[a]

POCID	IPOCID	POS
69.64%	43.28%	59.55%

[a]POCID, prediction of change in direction; IPOCID, independent prediction of change in direction; POS, prediction of sign.
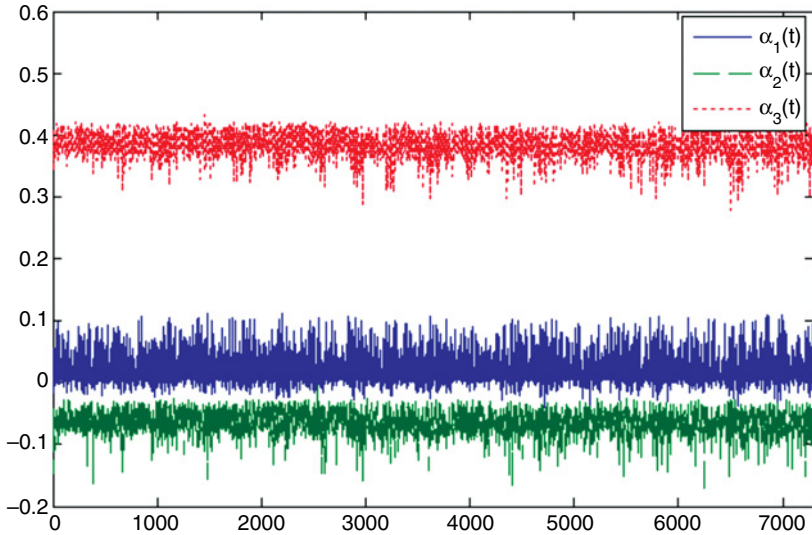
Figure 9.10   *Daily variations of the speed of mean reversion function $a_i(t)$ in New York.*

TABLE 9.13   Estimated Parameters of the Seasonal Variance for the WN

c_0	c_1	c_2	c_3	c_4	d_1	d_2	d_3	d_4
0.0935	−0.000020	−0.0034	0.0014	−0.0026	0.0353	−0.0016	−0.0042	−0.0052

case of the linear model. The distributional statistics of the residuals are presented in Table 9.14.

The distributional statistics of the residuals indicate that in-sample the two models can represent the dynamics of the DAWSs accurately; however, an improvement is evident when a nonlinear nonparametric wavelet network is used.

FORECASTING DAILY AVERAGE WIND SPEEDS

In this section the model proposed is validated out-of-sample. In addition, the performance of the model is tested against two models: first, against the linear model described previously, and second, against the simple persistent method, usually referred to as the benchmark. The linear model is the AR(4) model described in

TABLE 9.14   Descriptive Statistics of the Residuals for the Wavelet Network Model[a]

Var.	Mean	S. dev.	Max.	Min.	Skew	Kur.	JB	p-Value	KS	p-Value	LBQ	p-Value
Noise	0	1	3.32	−4.91	−0.08	3.04	8.84	0.0043	0.927	0.3544	13.437	0.858

S. dev., standard deviation; JB, Jarque–Bera statistic; KS, Komogorov–Smirnov statistic; LBQ, Ljung–Box Q-statistic.

the preceding section. The persistent method assumes that today's and tomorrow's DAWSs will be equal [i.e., $W^*(t + 1) = W(t)$, where W^* indicates the forecasted value].

The three models will be used for forecasting DAWSs for two different periods. Usually, wind derivatives are written for a period of a month. Hence, DAWSs for one and two months will be forecasted. The out-of-sample data set corresponds to the period from January 1 to February 28, 2008 and was not used for the estimation of the linear and nonlinear models. Note that our previous analysis reveals that the variance is higher in the winter period, indicating that it is more difficult to forecast DAWS accurately for these two months. To compare our results, the Monte Carlo approach is followed. We simulate 10,000 paths and calculate the average error criteria.

The performance of the three methods when the forecast window is one month is presented in Table 9.15. Various error criteria are estimated, such as the mean error, the median, Max. AE, the MSE, the POCID, and the IPOCID. As shown in the table, our proposed method outperforms both the persistent and the AR(4) model. The AR(4) model performs better than the naive persistent method; however, all error criteria are improved further when a nonlinear wavelet network is used. The MSE is 16.3848 for the persistent method, 10.5376 for the AR(4) model, and 10.4643 for the wavelet network. In addition, our model can predict the movements of the wind speed more accurately since the POCID is 80% for the wavelet network and the AR(4) models, whereas it is only 47% for the persistent method. Moreover, the IPOCID is 37% for the model proposed, whereas it is only 33% for the other two methods.

Next, the three forecasting methods are evaluated for two months of day-ahead forecasts. The results are similar and are presented in Table 9.16. The wavelet network proposed outperforms the other two methods. Only the Md.AE is slightly better when the AR(4) model is used. However, the IPOCID is 38% for the AR(4) method, whereas

TABLE 9.15 Out-of-Sample Comparison of One-Month Forecasts of DAWS[a]

	Persistent	AR(4)	WN
Md.AE	2.3000	2.3363	2.0468
ME	−0.0483	0.2117	−0.0485
MAE	3.3000	2.5403	2.5151
MaxAE	8.2000	7.9160	7.7019
SSE	507.9300	326.6666	324.3940
RMSE	4.0478	3.2461	3.2348
NMSE	1.5981	1.0278	1.0206
MSE	16.3848	10.5376	10.4643
MAPE	0.3456	0.2724	0.2680
SMAPE	0.3233	0.2555	0.2518
POCID	47%	80%	80%
IPOCID	33%	33%	37%
POS	100%	100%	100%

[a]Md.AE, median absolute error; ME, mean error; MAE, mean absolute error; MaxAE, maximum absolute error; SSE, sum of squared errors; RMSE, root mean squared error; NMSE, normalized mean squared error; MSE, mean squared error; MAPE, mean absolute percentage error; SMAPE, symmetric MAPE; POCID, position of change in direction; IPOCID, independent POCID; POS, position of sign.

TABLE 9.16   Out-of-Sample Comparison of Two-Month Forecasts of DAWS[a]

	Persistent	AR(4)	WN
Md.AE	2.4000	2.6393	2.6745
ME	0.1101	0.3570	0.1616
MAE	3.3678	2.8908	2.7967
MaxAE	11.2000	10.0054	8.3488
SSE	1054.3500	754.9589	688.2363
RMSE	4.2273	3.5771	3.4154
NMSE	1.4110	1.0103	0.9210
MSE	17.8703	12.7959	11.6650
MAPE	0.3611	0.3126	0.3056
SMAPE	0.3289	0.2808	0.2778
POCID	45%	69%	69%
IPOCID	36%	38%	43%
POS	100%	100%	100%

[a]Md.AE, median absolute error; ME, mean error; MAE, mean absolute error; MaxAE, maximum absolute error; SSE, sum of squared errors; RMSE, root mean squared error; NMSE, normalized mean squared error; MSE, mean squared error; MAPE, mean absolute percentage error; SMAPE, symmetric MAPE; POCID, position of change in direction; IPOCID, independent POCID; POS, position of sign.

it is 43% for the wavelet network. Also, our results indicate that the benchmark persistent method produces significantly poorer forecasts.

Our results indicate that the wavelet network can forecast evolution of the dynamics of the DAWSs and hence constitute an accurate tool for wind derivative pricing. The cumulative average wind speed (CAWS) index is calculated to provide better insight into the performance of each method. Since we are interested in weather derivatives, one common index is the sum of the daily average wind speed index over a specific period. An estimation of three methods is presented in Table 9.17. The wavelet network, the AR(4), and the historical burn analysis (HBA) method are compared. The HBA is a simple statistical method that estimates the performance of the index over a specific period in the past and is often used in the industry. It represents the average of 20 years of the index during January and February and serves as a benchmark.

The final row of Table 9.17 presents the actual values of the cumulative rainfall index. An inspection of the table reveals that the wavelet network significantly outperforms the other two methods. For the first case, where forecasts for one month ahead are estimated, the forecast of the CAWS index using the wavelet network is 312.7, while the actual index is 311.2. On the other hand, the forecast using the AR(4) model is 305.1. However, when the forecast period is increased, the forecast of the AR(4) model deviates significantly. For the second case, the forecast of the wavelet

TABLE 9.17   Estimation of the Cumulative Rainfall Index for 1 and 2 Months Using an AR(4) model, Wavelet Network, and Historical Burn Analysis

	AR(4)	WN	HBA	Actual
1 month	305.1	312.7	345.5	311.2
2 months	579.5	591.1	658.3	600.6

network is 591.1, the actual index is 600.6, and the AR(4) forecast is 579.5. Finally, we have to mention that the wavelet network uses less information than the AR(4) model, since with the wavelet network only the information of from three lags is used.

Since we are interested in wind derivatives and the valuation of wind contracts, an illustration of the performance of each method using a theoretical contract is presented next. A common wind contract has a tick size of 0.1 knot and pays $20 per tick size. Hence, for a one-month contract the AR(4) method underestimates the contract size by $1200, while the wavelet network only overestimates the contract by $300. Similarly, for a two-month contract the AR(4) method underestimates the contract size by $4220, while the wavelet network underestimates the contract by $1900.

Incorporating meteorological forecasts can lead to a potentially significant improvement in the performance of the model proposed. Meteorological forecasts can easily be incorporated in both the linear and wavelet network models presented previously. A similar approach was followed for temperature derivatives by Dorfleitner and Wimmer (2010). However, this method cannot always be applied. Despite great advances in meteorological science, weather still cannot be predicted precisely and consistently, and forecasts beyond 10 days are not considered accurate (Wilks, 2011). If the day that the contract is traded is during or close to the life of the derivative (during the period that wind measurements are considered), the meteorological forecasts can be incorporated in order to improve the performance of the methods. However, very often, weather derivatives are traded long before the start of the life of the derivative. More precisely, very often weather derivatives are traded months or even a season before the starting day of the contract. In that case, meteorological forecasts cannot be used.

CONCLUSIONS

In this chapter the DAWSs from New York were studied. Our analysis revealed strong seasonality in the mean and variance. The DAWSs were modeled by a mean reverting Ornstein–Uhlenbeck process in the context of wind derivative pricing. In this study the dynamics of the wind-generating process are modeled using a nonparametric nonlinear wavelet network. Our proposed methodology was compared in-sample and out-of-sample against two methods often used in prior studies. The characteristics of the wind speed process are very similar to the process of daily average temperatures. Our results indicate a slight downward trend and seasonality in the mean and variance. In addition, the seasonal variance is higher in winter, reaching its lower values during the summer period.

Our method is validated in a two-month-ahead out-sample forecast period. Moreover, the various error criteria produced by the wavelet network are compared against the linear AR model and the persistent method. Results show that the wavelet network outperforms the other two methods, indicating that wavelet networks constitute an accurate model for forecasting DAWSs. More precisely, the wavelet network forecasting ability is stronger in both samples. When we test the fitted residuals of the wavelet network, we observe that the distribution of the residuals is very close to

normal. Also, the wavelet network needed only the information of the past three days, while the linear method suggested a model with four lags. Finally, although we focused on DAWSs, our model can easily be adapted in hourly modeling.

The results in this case study are preliminary and can be analyzed further. More precisely, alternative methods for estimating the seasonality in the mean and in the variance can be developed. Alternative methods could improve the fitting to the original data as well as the training of the wavelet network. In addition, the inclusion of meteorological forecasts can further improve the forecasting performance of the wavelet networks.

It is also important to test the largest forecasting window of each method. Since meteorological forecasts of a window larger than a few days are considered inaccurate, this analysis will suggest the best model according to the forecasting interval desired.

Finally, a large-scale comparison must be conducted. Testing the methods proposed as well as more sophisticated models such as general ARFIMA or GARCH in various meteorological stations will provide a better insight into the dynamics of the DAWS as well as in the predictive ability of each method.

REFERENCES

Ailliot, P., Monbet, V., and Prevosto, M. (2006). "An autoregressive model with time-varying coefficients for wind fields." *Envirometrics*, 17, 107–117.

Alaton, P., Djehince, B., and Stillberg, D. (2002). "On modelling and pricing weather derivatives." *Applied Mathematical Finance*, 9, 1–20.

Alexiadis, M. C., Dokopoulos, P. S., Sahsamanoglou, H. S., and Manousaridis, I. M. (1998). "Short-term forecasting of wind speed and related electrical power." *Solar Energy*, 63(1), 61–68.

Barbounis, T. G., Theocharis, J. B., Alexiadis, M. C., and Dokopoulos, P. S. (2006). "Long-term wind speed and power forecasting using local recurrent neural network models." *IEEE Transactions on Energy Conversion,* 21(1), 273–284.

Benth, F. E., and Saltyte-Benth, J. (2009). "Dynamic pricing of wind futures." *Energy Economics*, 31, 16–24.

Beyer, H. G., Degner, T., Hausmann, J., Hoffmann, M., and Rujan, P. (1994). "Short-term prediction of wind speed and power outputof a wind turbine with neural networks." *2nd European Congress on Intelligent Techniques and Soft Computing*, Aachen, Germany.

Billinton, R., Chen, H., and Ghajar, R. (1996). "Time-series models for reliability evaluation of power systems including wind energy." *Microelectronics and Reliability*, 36(9), 1253–1261.

Brody, C. D., Syroka, J., and Zervos, M. (2002). "Dynamical pricing of weather derivatives." *Quantitave Finance*, 2, 189–198.

Brown, B. G., Katz, R. W., and Murphy, A. H. (1984). "Time-series models to simulate and forecast wind speed and wind power." *Journal of Climate and Applied Meteorology*, 23, 1184–1195.

Cao, M., and Wei, J. (2003). "Weather Derivatives: a New Class of Financial Instruments." Working Paper, University of Toronto, Toronto, Canada.

Caporin, M., and Pres, J. (2010). "Modelling and forecasting wind speed intensity for weather risk management." *Computational Statistics and Data Analysis*.

Castino, F., Festa, R., and Ratto, C. F. (1998). "Stochastic modelling of wind velocities time-series." *Journal of Wind Engineering and Industrial Aerodynamics*, 74–76, 141–151.

Celik, A. N. (2004). "A statistical analysis of wind power density based on the Weibull and Rayleigh models at the southern region of Turkey." *Renewable Energy*, 29(4), 593–604.

Cripps, E., Nott, D., Dunsmuir, W. T. M., and Wikle, C. (2005). "Space-time modelling of Sydney Harbour winds." *Australian and New Zealand Journal of Statistics*, 47(1), 3–17.

Daniel, A. R., and Chen, A. A. (1991). "Stochastic simulation and forecasting of hourly average wind speed sequences in Jamaica." *Solar Energy*, 46(1), 1–11.

Dorfleitner, G., and Wimmer, M. (2010). "The pricing of temperature futures at the Chicago Mercantile Exchange." *Journal of Banking and Finance*.

Dorvlo, A. S. S. (2002). "Estimating wind speed distribution." *Energy Conversion and Management*, 43(17), 2311–2318.

Garcia, A., Torres, J. L., Prieto, E., and de Francisco, A. (1998). "Fitting wind speed distributions: a case study." *Solar Energy*, 62(2), 139–144.

Haslett, J., and Raftery, A. E. (1989). "Space-time modelling with long-memory dependence: assessing Ireland's wind power resource." *Journal of the Royal Statistical Society,* Ser C, 38(1), 1–50.

Huang, Z., and Chalabi, Z. S. (1995). "Use of time-series analysis to model and forecast wind speed." *Journal of Wind Engineering and Industrial Aerodynamics*, 56, 311–322.

Jaramillo, O. A., and Borja, M. A. (2004). "Wind speed analysis in La Ventosa, Mexico: a bimodal probability distribution case." *Renewable Energy*, 29(10), 1613–1630.

Jewson, S., Brix, A., and Ziehmann, C. (2005). *Weather Derivative Valuation: The Meteorological, Statistical, Financial and Mathematical Foundations*, Cambridge University Press, Cambridge, UK.

Justus, C. G., Hargraves, W. R., Mikhail, A., and Graber, D. (1978). "Methods for estimating wind speed frequency distributions." *Journal of Applied Meteorology,* 17(3), 350–385.

Kamal, L., and Jafri, Y. Z. (1997). "Time-series models to simulate and forecast hourly averaged wind speed in Quetta, Pakistan." *Solar Energy*, 61(1), 23–32.

Kavak Akpinar, E., and Akpinar, S. (2005). "A statistical analysis of wind speed data used in installation of wind energy conversion systems." *Energy Conversion and Management*, 46(4), 515–532.

Kavasseri, R. G., and Seetharaman, K. (2009). "Day-ahead wind speed forecasting using f-ARIMA models." *Renewable Energy*, 34(5), 1388–1393.

Martin, M., Cremades, L. V., and Santabarbara, J. M. (1999). "Analysis and modelling of time-series of surface wind speed and direction." *International Journal of Climatology*, 19, 197–209.

Mohandes, M. A., Rehman, S., and Halawani, T. O. (1998). "A neural networks approach for wind speed prediction." *Renewable Energy*, 13(3), 345–354.

Mohandes, M. A., Halawani, T. O., Rehman, S., and Hussain, A. A. (2004). "Support vector machines for wind speed prediction." *Renewable Energy*, 29(6), 939–947.

More, A., and Deo, M. C. (2003). "Forecasting wind with neural networks." *Marine Structures*, 16, 35–49.

Nfaoui, H., Buret, J., and Sayigh, A. A. M. (1996). "Stochastic simulation of hourly average wind speed sequences in Tangiers (Morocco)." *Solar Energy*, 56(3), 301–314.

Nielsen, T. S., Madsen, H., Nielsen, H. A., Pinson, P., Kariniotakis, G., Siebert, N., Marti, I., Lange, M., Focken, U., Bremen, L. V., Louka, G., Kallos, G., and Galanis, G. (2006). "Short-term wind power forecasting using advanced statistical methods." *European Wind Energy Conference*, Athens, Greece.

Pinson, P., and Kariniotakis, G. N. (2003). "Wind power forecasting using fuzzy neural networks enhanced with on-line prediction risk assessment." *Power Tech Conference Proceedings IEEE Bologna*, Vol. 2.

Saltyte-Benth, J., and Benth, F. E. (2010). "Analysis and modelling of wind speed in New York." *Journal of Applied Statistics*, 37(6), 893–909.

Sfetsos, A. (2000). "A comparison of various forecasting techniques applied to mean hourly wind speed time-series." *Renewable Energy*, 21, 23–35.

Sfetsos, A. (2002). "A novel approach for the forecasting of mean hourly wind speed time series." *Renewable Energy*, 27, 163–174.

Tol, R. S. J. (1997). "Autoregressive conditional heteroscedasticity in daily wind speed measurements." *Theoretical and Applied Climatology*, 56, 113–122.

Tolman, H. L., and Booij, N. (1998). "Modeling wind waves using wavenumber–direction spectra and a variable wavenumber grid." *Global Atmosphere and Ocean System*, 6, 295–309.

Torres, J. L., Garcia, A., De Blas, M., and De Francisco, A. (2005). "Forecast of hourly average wind speed with ARMA models in Navarre (Spain)." *Solar Energy*, 79, 65–77.

Tuller, S. E., and Brett, A. C. (1984). "The characteristics of wind velocity that favor the fitting of a Weibull distribution in wind speed analysis." *Journal of Climate and Applied Meteorology*, 23(1), 124–134.

Wilks, D. S. (2011). *Statistical Methods in the Atmospheric Sciences*. Academic Press, Oxford, UK.

WRMA. (2010). "Weather derivatives volume plummets." Retrieved January, 2010, from www.wrma.org/pdf/weatherderivativesvolumeplummets.pdf

Yamada, Y. (2008). "Simultaneous optimization for wind derivatives based on prediction errors." *Americal Control Conference*, Washington, DC, 350–355.

Zapranis, A., and Alexandridis, A. (2008). "Modelling temperature time dependent speed of mean reversion in the context of weather derivetive pricing." *Applied Mathematical Finance*, 15(4), 355–386.