Available online at www.sciencedirect.com
SCIENCE DIRECT®
Solar Energy 79 (2005) 65–77
SOLAR ENERGY
www.elsevier.com/locate/solener

Forecast of hourly average wind speed with ARMA models in Navarre (Spain)

J.L. Torres [a,*], A. García [a], M. De Blas [a], A. De Francisco [b]

[a] *Department of Projects and Rural Engineering, Public University of Navarre, Edificio Los Olivos, 31006 Pamplona, Spain*
[b] *Department of Forestry Engineering, Polytechnic University of Madrid, Ciudad Universitaria, 28040 Madrid, Spain*

Received 4 February 2004; received in revised form 28 September 2004; accepted 28 September 2004
Available online 3 November 2004

Communicated by: Associate Editor Cornelis P. van Dam

Abstract

In this article we have used the ARMA (autoregressive moving average process) and persistence models to predict the hourly average wind speed up to 10 h in advance. In order to adjust the time series to the ARMA models, it has been necessary to carry out their transformation and standardization, given the non-Gaussian nature of the hourly wind speed distribution and the non-stationary nature of its daily evolution. In order to avoid seasonality problems we have adjusted a different model to each calendar month. The study expands to five locations with different topographic characteristics and to nine years. It has been proven that the transformation and standardization of the original series allow the use of ARMA models and these behave significantly better in the forecast than the persistence model, especially in the longer-term forecasts. When the acceptable RMSE (root mean square error) in the forecast is limited to 1.5 m/s, the models are only valid in the short term.

Keywords: Time series; Wind speed; Forecasting

1. Introduction

The forecast of hourly average wind speed 1–10 h in advance, and thereby the power output of wind farm, is of interest for the operation of conventional electric power plants that are connected to the same power grid as those conversion systems (Geerts, 1984; Sfestos, 2000). This is particularly important in weak grids. The modeling and prediction of time series of hourly average wind speed has been a subject of attention to a large number of researchers. The initial studies carried out used a Monte Carlo method to generate simulations when the parameters of the wind speed distribution were known. The series obtained using this method presented the weakness of not considering the autocorrelation of the hourly average wind velocities. Later, Chou and Corotis (1981) included the effect of autocorrelation, but they did not consider the non-Gaussian nature of the wind speed distribution. Brown et al. (1984) proposed a method that took into account the autocorrelated nature, the daytime non-seasonality, and the non-Gaussian shape of the wind speed distribution, applying a pure autoregressive model (AR) to a series

* Corresponding author. Tel.: +34 948 169 175; fax: +34 948 169 148.
E-mail address: jlte@unavarra.es (J.L. Torres).

0038-092X/$ - see front matter
doi:10.1016/j.solener.2004.09.013

Nomenclature

χ^2	statistic distribution chi-square (dimensionless)
ϕ_i	autoregressive parameters (dimensionless)
θ_j	parameters of moving average (dimensionless)
$\mu(t)$, $\sigma(t)$	mean and standard deviation periodic functions (m/s)
$a_{n,y}$	white noise variables (m/s)
AIC	Akaike Information Criterion
ARMA	Autoregressive Moving Average
B	delay operator
BIC	Bayesian Information Criterion
C	scale factor of Weibull distribution (m/s)
c_k	autocovariance coefficients (m/s)
d	number of days of the month
$\varphi_{n,n}$	partial autocorrelations coefficient (dimensionless)
K	shape factor of the Weibull distribution (dimensionless)
L	maximum delay considered (h)
m	exponent of transformation
M	mean of the $M_{n,y}$ series
MABE	mean absolute error (m/s)
$M_{n,y}$	transformed value of wind speed
n	subindex indicating the hour of a registered datum in a given month, i.e. $n \in [1,744]$ for a 31 days month
N	number of hours of the month
p	order of the autoregressive process
q	order of the moving average process
r_k	autocorrelation coefficient (dimensionless)
RMSE	root mean square error (m/s)
s	standard deviation of the series (m/s)
S_m	skewness statistic (dimensionless)
T	total number of parameters estimated
$V_{n,y}$	speed of the time series of a given month for the year y (m/s)
Y	number of years considered
y	subindex indicating the year of a registered datum in the series time

of data observed in one month. They pointed out that it would be more appropriate to use wind speed data from several years for each given month, despite the fact that the process would become more cumbersome and the formulas used would have to be modified. Geerts (1984) used an ARMA model for a single series one year long with the same goal of forecasting wind speed values in a relatively short term. The author compared the results with those obtained using a persistence model, concluding that for longer-term predictions than 1 h ARMA worked better than the persistence model. Notwithstanding, the RMSE for 10 h in advance exceeded in both cases the limit of 1.5 m/s that he established as acceptable. Balouktsis et al. (1986) applied the ARMA models to the wind speed time series from three locations with one- and two-year long records, but the prior transformation of the observed data was different from the one proposed by Brown, pointing out that the results were satisfactory. Daniel and Chen (1991) applied the ARMA model to three year long time series, and specifically for three months. For doing so, they admitted that the model used to generate time series of hourly average wind speed based on data from several years could provide future wind velocities more representative than those from the model based on one month data. Cited authors did forecasts 1–6 h in advance and they observed the deterioration of the results when the forecast was predicted more than 2 h in advance. Following the same procedure, Nfaoui et al. (1996) concluded that an AR (2) model is capable of simulating well the wind speed series recorded, that in this case were referred to only one location and had a time span of 12 years, and Kamal and Jafri (1997) confirmed that such procedure is useful to predict past values as well as the forecasted wind data of Quetta (Pakistan).

The aim of this work is to evaluate the applicability of the ARMA models (see Appendix A) to the time series of hourly average wind speed, and assess the predictive behaviour of the models obtained. The application of ARMA models requires time series to be stationary, i.e. the method assumes that the process remains in equilibrium about a constant mean level. The analysis was done from data collected in five weather stations located in two distinct areas of Navarre, one with smooth topography and the other in a mountainous region.

2. Materials and methods

We collected data from fourteen automated weather stations distributed across the entire territory of the Regional Community of Navarre (Spain) that had long enough data series without gaps. Among other meteorological variables, these stations record wind speed at a height of 10 m every ten minutes with cup anemometers. The value of the hourly average wind speed was obtained averaging the six values measured within each hour.

Given the fact that nine out of the fourteen stations showed average annual hourly velocities that were too low as to be of interest for wind power generation, we have restricted the study to just five stations. Two of these (Trinidad and Aralar) are located in a mountainous area, whereas (Lomanegra, Ujue and Yugo) are located in an open area with smooth topography. Table 1 lists the most relevant data of the historical series for these stations. In addition, as in previous studies, already referred in the introduction above, in order to avoid possible seasonality problems, we processed the data on a monthly basis, so that for each calendar month, the series being analyzed is made up by the records from the same month in the different years considered in each case by the data collection campaign.

2.1. Transformation and standardization of the observed data series

Hourly wind speed distributions in Navarre are better adjusted to a Weibull than to a Normal distribution (García et al., 1998; Torres et al., 1999). In addition, it is considered that the evolution of the hourly wind speed during the day is not stationary but it generally shows a cyclic behaviour, due among other reasons to atmospheric stability and instability phenomena. For these reasons, in order to apply the ARMA model we firstly have to transform and standardize the data of the observed monthly series.

This transformation is carried out by raising each one of the observed hourly values to the same index m, so that the distribution becomes approximately Gaussian. This is based on the fact that a Weibull variable raised to an index continues to be a Weibull variable (Eq. (1a)).

Given that Dubey (1967) demonstrated that with a form factor (K) close to 3.6 the Weibull distribution is similar to Normal, in order to do a preliminary estimation of the index to the power of which we must raise each one of the elements of the time series considered,

Table 1
Annual and total number of observations (N), average speeds (V) and standard deviations (σ) of the year series obtained from the stations

		Aralar	Trinidad	Lomanegra	Ujué	Yugo
1992	V_{1992} (m/s)	–	7.28	6.99	6.06	5.60
	N_{1992}	–	8091	7220	8432	8425
	σ_{1992}	–	3.61	4.66	3.15	2.92
1993	V_{1993} (m/s)	–	6.73	6.97	5.97	5.38
	N_{1993}	–	8592	8506	8497	8461
	σ_{1993}	–	3.58	4.40	3.00	2.84
1994	V_{1994} (m/s)	7.31	7.39	7.33	6.17	5.72
	N_{1994}	8637	8632	8473	8206	7781
	σ_{1994}	3.42	3.41	4.40	2.99	2.82
1995	V_{1995} (m/s)	6.85	8.27	8.39	6.05	5.87
	N_{1995}	8649	8008	8413	8471	6373
	σ_{1995}	3.76	4.55	5.18	3.43	2.69
1996	V_{1996} (m/s)	7.79	7.52	7.86	6.66	5.95
	N_{1996}	8591	8454	8517	8567	8009
	σ_{1996}	4.56	3.55	4.53	3.20	2.82
1997	V_{1997} (m/s)	7.51	6.16	5.57	5.60	5.53
	N_{1997}	3665	3713	2415	3659	5140
	σ_{1997}	3.60	2.93	3.38	2.84	3.77
1998	V_{1998} (m/s)	7.66	8.94	7.60	6.47	5.66
	N_{1998}	5074	2953	5259	5800	5772
	σ_{1998}	3.59	8.66	4.47	3.13	2.85
1999	V_{1999} (m/s)	–	6.97	7.72	5.87	5.28
	N_{1999}	–	3431	1056	1055	1053
	σ_{1999}	–	3.62	4.64	3.28	2.84
2000	V_{2000} (m/s)	3.67	7.05	7.30	6.41	5.60
	N_{2000}	670	5850	5282	4966	4958
	σ_{2000}	4.03	3.28	4.40	3.25	2.79
Totals	V_{media}	7.32	7.36	7.43	6.19	5.66
	N_{Total}	35,286	57,724	55,141	57,653	55,972
	σ_{Total}	3.91	4.11	4.60	3.16	2.93

we have calculated the value of m from the following expression (Eq. (1b)):

$$P(v) = \frac{K}{C}\left(\frac{V}{C}\right)^{K-1} \mathrm{e}^{-\left(\frac{V}{C}\right)^{K}}, \tag{1a}$$

$$m = K/3.6, \tag{1b}$$

where K is the form factor of the Weibull distribution of observed hourly wind speed.

According to what Daniel and Chen (1991) proposed, also used later on by Nfaoui et al. (1996), the previous value (m) has been used as reference to obtain more accurately the index of the transformation by another alternative method based on the statistic that evaluates the symmetry of the distribution. This is given by the expression:

$$S_m = \sum_{y=1}^{Y} \sum_{n=1}^{N} \frac{\left[(M_{n,y} - \overline{M})/s\right]^3}{Y \cdot N}, \tag{2}$$

where $M_{n,y}$ is the transformed value of wind speed; $\overline{M}$ the mean of $M_{n,y}$; s the standard deviation of the series; being Y the number of years considered, N the number of hours of the month in question, and the subscripts n and y representing respectively the time of the month and the year.

This method requires an iterative calculation in which the original series is raised to the power of different indices and the corresponding S_m are computed, finally choosing the one that makes the value of S_m closest to zero. The value of m obtained in the preliminary estimation mentioned before can be used as an initial reference value for the iterations. The m values computed for the different months in each season are listed in Table 2 (column denoted by m asymmetry).

By the end of this stage, each speed $V_{n,y}$ of the actual time series of a given month is transformed into another variable $V'_{n,y}$. In Fig. 1 we have included as examples the monthly charts of the hourly average values for the months of January and July of the transformed wind speed series obtained from the data collected at the Trinidad station. The wind variation throughout the average day can be noticed in these charts.

In an effort to eliminate the daily seasonality that may occur, as revealed by Fig. 1, the next step is the standardization of each one of the time series. For doing so, as shown in Eq. (3), the expected data value in an hour, i.e. the average of all the data of the transformed series from the same time, is substracted from each $V'_{n,y}$ data. Result is divided by the hourly standard deviation. This way, for each month of the year and each weather station we obtain a series of transformed and standardized values $V^*_{n,y}$:

$$V^*_{n,y} = \frac{V'_{n,y} - \mu(t)}{\sigma(t)}, \tag{3}$$

with

$$\mu(t) = \frac{\sum_{i=0}^{d\cdot Y-1} V'_{24\cdot i+t}}{d \cdot Y}, \quad 1 \leqslant t \leqslant 24,$$

$$\sigma(t) = \left[\frac{\sum_{i=0}^{d\cdot Y-1} (V'_{24\cdot i+t} - \mu(t))^2}{d \cdot Y}\right]^{1/2}, \quad 1 \leqslant t \leqslant 24,$$

and d being the number of days of the month considered.

It is assumed that both $\mu(t)$ and $\sigma(t)$ are periodic functions. Therefore, for example, $\mu(1)$ and $\sigma(1)$ are used in the standardization of $V'_1, V'_{25}, \ldots$ or $\mu(2)$ and $\sigma(2)$ in that of V'_2, V'_{26}.

2.2. Process of the ARMA model

Once the data series have been adequately transformed and standardized, we can begin the process of construction of the model that will allow carrying out the forecasts. In this study we will use the ARMA model that, as we mentioned above, its use for the forecast of the behaviour of average wind speed is well documented by Daniel and Chen (1991) and Kamal and Jafri (1997).

Basically, in the ARMA models the forecast of the wind speed depends not only on the values it has had in the more or less recent past according to the autoregressive component, but it can also be a function of the residuals of past forecasts, that correspond to previous hours to that for what we are doing the forecast. The mathematical expression of the general ARMA model (p, q) that is applied in this case to the series of transformed and standardized values is the following equation:

$$\begin{aligned}(1 - \phi_1 \cdot B - \phi_2 \cdot B^2 - \cdots - \phi_p \cdot B^p) \cdot V^*_{n,y} \\ = (1 - \theta_1 \cdot B - \theta_2 \cdot B^2 - \cdots - \theta_q \cdot B^q) \cdot a_{n,y},\end{aligned} \tag{4}$$

where B is a delay operator so that $B \cdot V^*_{n,y} = V^*_{n-1,y}$, $\phi_1, \ldots, \phi_p$ are the autoregressive parameters, $\theta_1, \ldots, \theta_q$ the parameters of moving average and $a_{n,y}$ a white noise (random uncorrelated variables with average value zero and variance σ_a^2).

The construction of the models of each of the monthly time series in the different weather stations studied, consists mainly in identifying the p and q indices of the models, determining the ϕ and θ parameters contained in them, and finally validating them. We have briefly outlined the procedures followed to put into practice the different phases that were necessary in constructing the model.

2.2.1. Identification phase

The main goal of this phase is determining the indices p and q of the model (in both cases they can have null values, in the assumption that we have mobile average or pure autoregressive models), and for that matter we

Table 2
Characteristics of the wind speed distributions and of the ARMA models selected for each month and weather station

Month		K	C, m/s	m Dubey	m asymmetry	ARMA model selected	Parameters		
							AR	MA	
January	Aralar	1.46	8.62	0.41	0.49	(1,2)	$\phi_1 = 0.9856$	$\theta_1 = -0.0390$	$\theta_2 = 0.0918$
	Trinidad	2.24	8.78	0.62	0.75	(1,1)	$\phi_1 = 0.9195$	$\theta_1 = -0.0974$	
	Lomanegra	2.10	10.11	0.58	0.60	(1,1)	$\phi_1 = 0.9273$	$\theta_1 = -0.1181$	
	Yugo	1.55	5.55	0.43	0.40	(1,1)	$\phi_1 = 0.9449$	$\theta_1 = -0.0126$	
	Ujue	2.45	8.58	0.68	0.68	(1,1)	$\phi_1 = 0.8871$	$\theta_1 = -0.1541$	
February	Aralar	2.04	9.50	0.57	0.68	(1,2)	$\phi_1 = 0.9289$	$\theta_1 = -0.0757$	$\theta_2 = 0.0546$
	Trinidad	2.21	9.42	0.61	0.54	(1,2)	$\phi_1 = 0.9309$	$\theta_1 = -0.1080$	$\theta_2 = 0.0527$
	Lomanegra	1.41	8.38	0.39	0.39	(1,4)	$\phi_1 = 0.9764$	$\theta_1 = -0.0477$ $\theta_2 = 0.0748$	$\theta_3 = 0.0434$ $\theta_4 = 0.0759$
	Yugo	1.74	6.26	0.48	0.50	(2,2)	$\phi_1 = 0.6059$ $\phi_2 = -0.6156$	$\theta_1 = 0.6445$	$\theta_2 = 0.1234$
	Ujue	1.90	7.74	0.53	0.50	(1,3)	$\phi_1 = 0.9318$	$\theta_1 = -0.0469$ $\theta_2 = 0.0902$	$\theta_3 = 0.0759$
March	Aralar	1.62	7.12	0.45	0.54	(1,3)	$\phi_1 = 0.9522$	$\theta_1 = -0.0501$ $\theta_2 = 0.1498$	$\theta_3 = 0.0746$
	Trinidad	2.17	8.50	0.60	0.72	(1,1)	$\phi_1 = 0.9408$	$\theta_1 = -0.0922$	
	Lomanegra	1.89	9.57	0.53	0.63	(1,4)	$\phi_1 = 0.9536$	$\theta_1 = -0.1000$ $\theta_2 = 0.0392$	$\theta_3 = -0.0125$ $\theta_4 = -0.0629$
	Yugo	2.16	6.68	0.60	0.72	(1,2)	$\phi_1 = 0.9431$	$\theta_1 = -0.0249$	$\theta_2 = 0.0855$
	Ujue	2.09	7.32	0.58	0.68	(1,2)	$\phi_1 = 0.9209$	$\theta_1 = -0.0208$	$\theta_2 = 0.1109$
April	Aralar	1.88	7.87	0.52	0.63	(1,8)	$\phi_1 = 0.9786$	$\theta_1 = -\ 0.0445$ $\theta_2 = 0.0704$ $\theta_3 = -\ 0.0088$ $\theta_4 = 0.0879$	$\theta_5 = 0.0774$ $\theta_6 = 0.0487$ $\theta_7 = 0.0474$ $\theta_8 = 0.0697$
	Trinidad	2.14	8.66	0.59	0.68	(1,1)	$\phi_1 = 0.936$	$\theta_1 = -0.0783$	
	Lomanegra	1.67	8.71	0.46	0.41	(1,0)	$\phi_1 = 0.9563$		
	Yugo	2.17	7.11	0.60	0.68	(1,2)	$\phi_1 = 0.9252$	$\theta_1 = -0.0247$	$\theta_2 = 0.1003$
	Ujue	2.05	7.46	0.57	0.57	(1,2)	$\phi_1 = 0.9099$	$\theta_1 = -0.0355$	$\theta_2 = 0.0797$
May	Aralar	0.20	7.47	0.56	0.67	(2,1)	$\phi_1 = 0.1952$ $\phi_2 = 0.6691$	$\theta_1 = -0.7910$	
	Trinidad	2.51	7.42	0.70	0.61	(1,1)	$\phi_1 = 0.9104$	$\theta_1 = -0.0921$	
	Lomanegra	1.68	8.30	0.47	0.41	(1,2)	$\phi_1 = 0.9502$	$\theta_1 = -0.0484$	$\theta_2 = 0.0829$
	Yugo	2.36	6.40	0.65	0.57	Not found			
	Ujue	2.42	7.07	0.67	0.62	(1,1)	$\phi_1 = 0.8574$	$\theta_1 = -0.0436$	
June	Aralar	1.85	8.32	0.51	0.45	Not found			
	Trinidad	2.38	7.90	0.66	0.69	Not found			
	Lomanegra	1.83	8.50	0.51	0.44	Not found			
	Yugo	2.20	6.45	0.61	0.53	(1,4)	$\phi_1 = 0.9449$	$\theta_1 = 0.0286$ $\theta_2 = 0.1296$	$\theta_3 = 0.0681$ $\theta_4 = 0.0715$
	Ujue	2.26	7.02	0.63	0.58	Not found			
July	Aralar	1.55	7.63	0.43	0.38	(1,1)	$\phi_1 = 0.9059$	$\theta_1 = -0.1767$	
	Trinidad	2.37	7.76	0.66	0.65	(1,1)	$\phi_1 = 0.9004$	$\theta_1 = -0.1855$	
	Lomanegra	1.81	7.55	0.50	0.44	(1,4)	$\phi_1 = 0.9460$	$\theta_1 = -0.0638$ $\theta_2 = 0.0589$	$\theta_3 = 0.0830$ $\theta_4 = 0.0724$
	Yugo	2.37	6.57	0.66	0.66	(1,2)	$\phi_1 = 0.8888$	$\theta_1 = -0.0921$	$\theta_2 = 0.0812$
	Ujue	2.34	6.72	0.65	0.57	(1,2)	$\phi_1 = 0.8441$	$\theta_1 = -0.0735$	$\theta_2 = 0.0858$
August	Aralar	2.54	6.79	0.70	0.85	(2,2)	$\phi_1 = 1.688$ $\phi_2 = -0.7003$	$\theta_1 = 0.6282$	$\theta_2 = 0.2236$
	Trinidad	1.81	6.90	0.50	0.60	(1,7)	$\phi_1 = 0.9823$	$\theta_1 = -0.0880$ $\theta_2 = 0.0954$ $\theta_3 = 0.0803$ $\theta_4 = 0.0971$	$\theta_5 = 0.0631$ $\theta_6 = 0.0390$ $\theta_7 = 0.0617$

(*continued on next page*)

Table 2 (*continued*)

Month		K	C, m/s	m Dubey	m asymmetry	ARMA model selected	Parameters		
							AR	MA	
	Lomanegra	1.17	8.11	0.47	0.41	(1,2)	$\theta_1 = 0.9609$	$\theta_1 = -0.0676$	$\theta_2 = 0.1670$
	Yugo	2.14	6.07	0.60	0.52	(1,1)	$\phi_1 = 0.8650$	$\theta_1 = -0.1144$	
	Ujue	1.45	5.07	0.40	0.48	Not found			
September	Aralar	2.48	7.97	0.69	0.60	(1,1)	$\phi_1 = 0.8776$	$\theta_1 = -0.1556$	
	Trinidad	2.51	8.21	0.70	0.68	(1,1)	$\phi_1 = 0.9008$	$\theta_1 = -0.1367$	
	Lomanegra	1.91	8.53	0.53	0.48	(1,1)	$\phi_1 = 0.9338$	$\theta_1 = -0.1486$	
	Yugo	2.07	5.84	0.58	0.50	(1,2)	$\phi_1 = 0.9040$	$\theta_1 = -0.0297$	$\theta_2 = 0.1285$
	Ujue	2.07	6.66	0.57	0.69	(1,2)	$\phi_1 = 0.9214$	$\theta_1 = -0.0179$	$\theta_2 = 0.1097$
October	Aralar	2.27	7.89	0.63	0.62	(1,3)	$\phi_1 = 0.9393$	$\theta_1 = -0.0930$	$\theta_3 = 0.0645$
								$\theta_2 = 0.1002$	
	Trinidad	2.10	7.71	0.58	0.51	(1,3)	$\phi_1 = 0.9538$	$\theta_1 = -0.0466$	$\theta_3 = 0.0583$
								$\theta_2 = 0.1214$	
	Lomanegra	1.70	8.26	0.47	0.41	(1,1)	$\theta_1 = 0.9549$	$\theta_1 = -0.0927$	
	Yugo	1.95	5.70	0.54	0.47	(1,2)	$\phi_1 = 0.9310$	$\theta_1 = 0.0125$	$\theta_2 = 0.1184$
	Ujue	2.19	6.98	0.61	0.61	(1,2)	$\phi_1 = 0.9359$	$\theta_1 = 0.0074$	$\theta_2 = 0.0993$
November	Aralar	2.24	9.48	0.62	0.59	(1,4)	$\phi_1 = 0.9448$	$\theta_1 = -0.1019$	$\theta_3 = 0.0154$
								$\theta_2 = 0.1153$	$\theta_4 = 0.0656$
	Trinidad	2.17	8.12	0.60	0.53	(1,2)	$\theta_1 = 0.9329$	$\theta_1 = -0.0721$	$\theta_2 = 0.1043$
	Lomanegra	1.74	8.48	0.48	0.43	(1,0)	$\phi_1 = 0.9530$		
	Yugo	1.72	4.75	0.48	0.50	(1,2)	$\phi_1 = 0.9291$	$\theta_1 = -0.0702$	$\theta_2 = 0.0679$
	Ujue	1.89	6.62	0.53	0.53	(1,2)	$\phi_1 = 0.9327$	$\theta_1 = 0.0037$	$\theta_2 = 0.1029$
December	Aralar	2.39	9.24	0.66	0.80	(1,2)	$\phi_1 = 0.9299$	$\theta_1 = -0.0861$	$\theta_2 = 0.0946$
	Trinidad	2.14	9.10	0.59	0.52	(1,5)	$\phi_1 = 0.9537$	$\theta_1 = -0.0646$	$\theta_4 = 0.0605$
								$\theta_2 = 0.0906$	$\theta_5 = 0.0533$
								$\theta_3 = 0.0234$	
	Lomanegra	1.58	8.20	0.44	0.39	(1,3)	$\phi_1 = 0.9615$	$\theta_1 = -0.0728$	$\theta_3 = 0.0620$
								$\theta_2 = 0.0746$	
	Yugo	1.73	5.83	0.48	0.42	(1,1)	$\phi_1 = 0.9155$	$\theta_1 = -0.1078$	
	Ujue	1.89	6.73	0.52	0.46	(1,2)	$\phi_1 = 0.9379$	$\theta_1 = 0.0265$	$\theta_2 = 0.1165$

initially calculate the autocorrelation (r_k) and partial autocorrelation coefficients that allow to come up with the corresponding autocorrelograms and partial autocorrelograms. The expression used for the computation of the autocorrelation coefficient with delay k, as proposed by Daniel and Chen (1991) is the following equation:

$$r_k = c_k/c_0,$$
$$c_k = \frac{1}{Y \cdot N - Y \cdot k}\left[\sum_{y=1}^{Y}\sum_{n=1}^{N-k}(V^*_{n,y} - \overline{V}^*)(V^*_{n+k,y} - \overline{V}^*)\right],$$
$$\overline{V}^* = \frac{1}{Y \cdot N}\left(\sum_{y=1}^{Y}\sum_{n=1}^{N} V^*_{n,y}\right), \tag{5}$$

where we take into consideration the fact that the time series that is analyzed in each case is formed by the transformed and standardized values of the records of the same month of successive years.

On the other hand, for the computation of the partial autocorrelation coefficients ($\varphi_{n,n}$) we have used the Durbins relations (see Appendix B). As an example, Fig. 2 includes the autocorrelogram and partial correlogram for the months of January and July from the Trinidad station.

The observation of the tendencies of the aforementioned autocorrelograms, the assessment of the corresponding autocorrelation coefficients with the appropriate acceptance statistics and, additionally, the application of the identification method proposed by Tsay and Tiao (1984) and Beguin et al. (1980), allows to identify an adequate group of models for each case. The final decision on the best one to choose is taken after results discussion obtained in the following estimation phase. This phase provides essential data for the application of decision criteria, such as the BIC (Bayesian Information Criterion) or the AIC (Akaike Information Criterion) that respond to the following expressions, respectively:

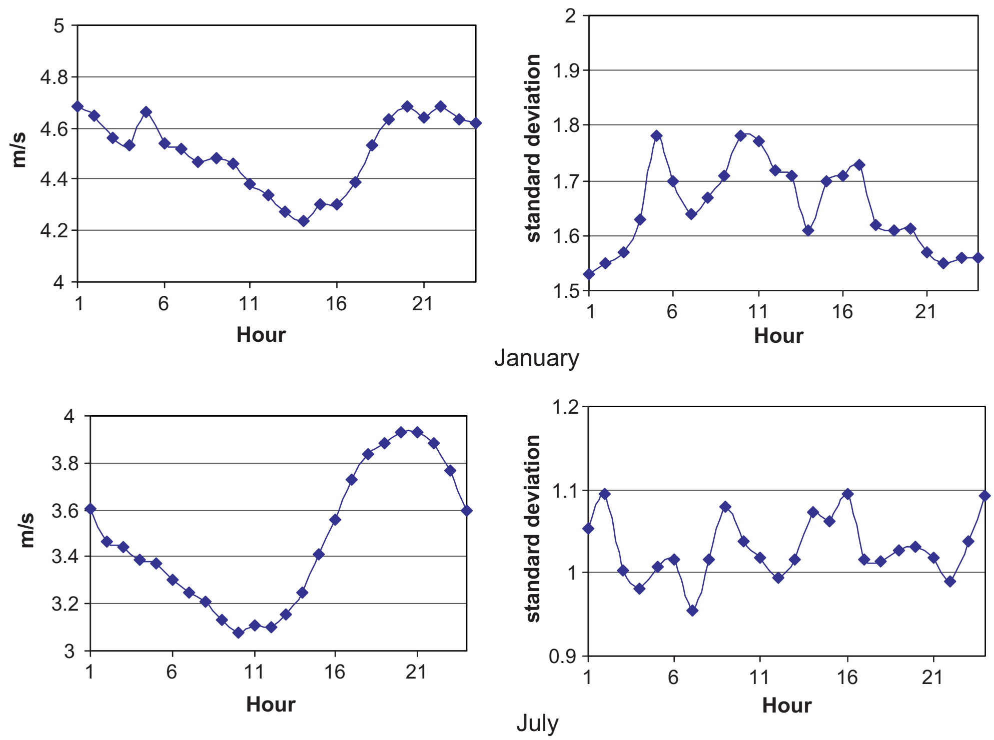


Fig. 1. Average hourly speeds (left) and standard deviations (right) of the transformed series of the months of January and July from the Trinidad station (Unit: m/s).

$$\begin{aligned} \mathrm{BIC}(p,q) &= (Y\cdot N)\cdot \ln(\sigma_a^2(p,q)) + T\cdot \ln(Y\cdot N), \\ \mathrm{AIC}(p,q) &= (Y\cdot N)\cdot \ln(\sigma_a^2(p,q)) + 2T, \end{aligned} \tag{6}$$

where T equals the total number of parameters to be estimated.

2.2.2. Parameter estimation phase

As it can be drawn from the preceding comments, this phase overlaps with the previous one in some aspects, and it is covered in two stages: in the first one there is a preliminary estimation of the values of the parameters included in the model being considered (given the processing that the observed data have undergone, the global constant that appears as a parameter in many of the ARMA models is set to zero) and in the second we use those values as input for a more accurate estimation.

The preliminary estimation is done applying the Yule-Walker relations for the autoregression coefficients, while the moving average coefficients are obtained by previously computing the corrected autocovariances and applying afterwards a Newton–Raphson algorithm, as proposed by Box and Jenkins (1976).

For the final estimation of the parameters we begin by doing a forecast backwards, with the aim that the focus of the estimation is not conditioned, and the parameters, both autoregressive and moving average are determined by minimizing the sum of the squares of the residuals generated, by means of the algorithm proposed by Marquardt. The final value of the variance of the white noise and the correlation matrix of the parameters are also determined. The value of the white noise variance is one of the data required to apply the decision criteria that we have referred to before.

2.2.3. Validation phase

Lastly, in order to test the model selected with its own parameters, we analyze the correlogram of the residuals obtained, an example of which can be seen in Fig. 3. A global contrast is applied using the statistic proposed by Box–Pierce (Eq. (7)), given by

$$Q = Y\cdot N\sum_{k=1}^{L} r_k^2(a), \tag{7}$$

where $r_k(a)$ is the autocorrelation coefficient of the residuals and L is the maximum delay considered.

In order to accept the model, we must prove that this statistic follows a χ^2 distribution, of as many degrees of freedom as L minus the number of parameters to be estimated, $(p + q)$.

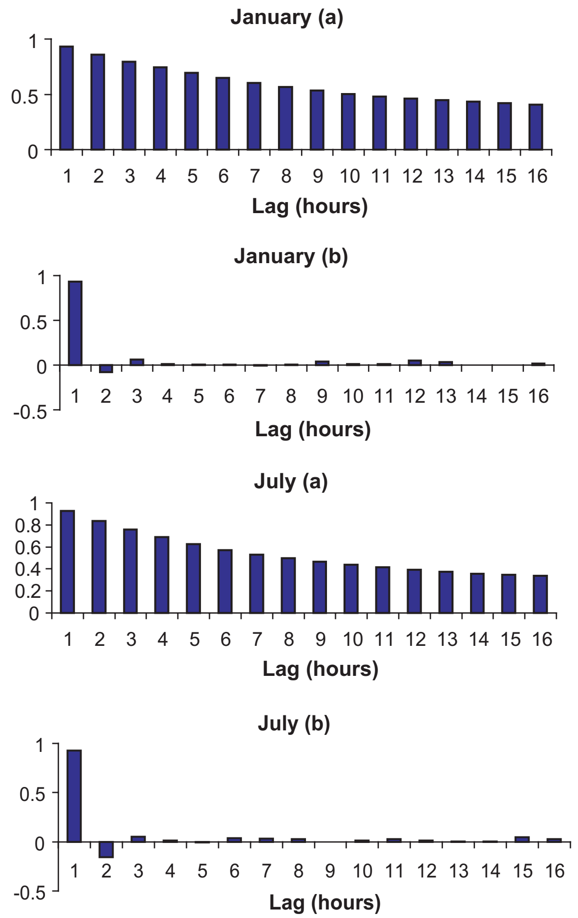


Fig. 2. Autocorrelation (a) and partial autocorrelation (b) coefficients of the time series of standardized wind speeds of the months of January and July from the Trinidad station, extended to a delay of 16 h.

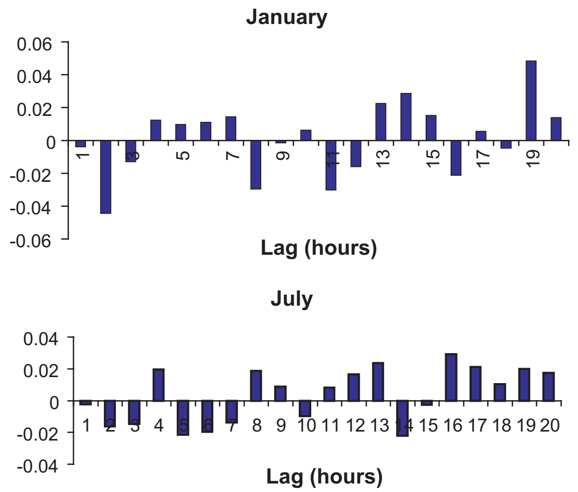


Fig. 3. Autocorrelograms of the residuals obtained by applying the models selected in the months indicated at the Trinidad station.

where k is the number of advance intervals of the forecast done at a time t.

By doing so, those values of V^*, on the right side of the equation, that are unknown, can be replaced by their respective forecasts, and their residuals, that relate to the moment from which the forecast is being done or a later one, are substituted by the expected value of zero.

Obviously, since the forecast is done on a transformed and standardized series, in order to obtain the forecasts of wind speed a last step of undoing de standardization and transformation is necessary. This way the forecasts can be compared with the actual values and assess their degree of success.

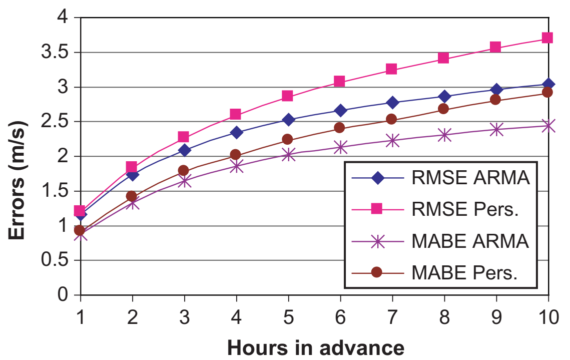


Fig. 4. RMSE and MABE of ARMA and persistence model based wind speed predictions at the Trinidad station in the month of July for forecast periods ranging from 1 to 10 h.

2.3. Process of the persistence model

The persistence model is a simple model that meets the following definition equation:

$$V^*_{n,y} = V^*_{n-k,y}, \tag{8}$$

where the subscript k represents the lag ($k = 1, 2, 3, \ldots$ hours).

2.4. Forecast

2.4.1. With ARMA models

Once the model has been built and validated for each month in each weather station, we can move on to the forecast phase. For this purpose, the definition equation of the ARMA model (Eq. (9)) is used:

$$V^*_{t+k} = \phi_1 \cdot V^*_{t+k-1} + \phi_2 \cdot V^*_{t+k-2} + \cdots + \phi_p \cdot V^*_{t+k-p} + a_{t+k} - \theta_1 \cdot a_{t+k-1} - \cdots - \theta_q \cdot a_{t+k-q}, \tag{9}$$

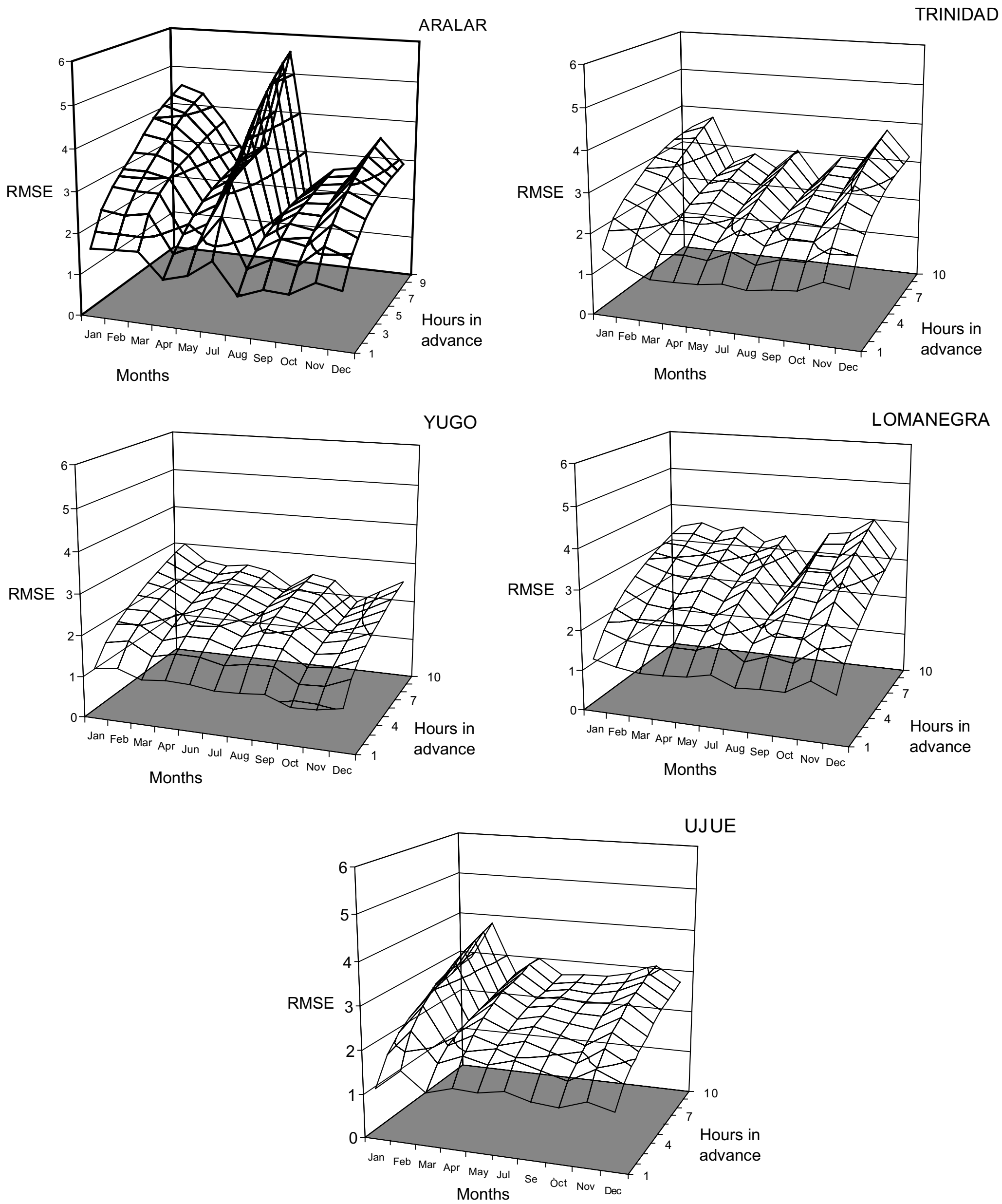


Fig. 5. Evolution of the RMSE (m/s) in the various months studied for different stations with forecasts of up to 10 h in advance.

2.4.2. With persistence model

The forecast using this model is done applying the expression Eq. (8), and same as with the ARMA model, it is updated as new observed values are known.

In order to assess the goodness of the forecast, both with ARMA models and with persistence models we have computed the MABE and RMSE between the forecasted series and the actual one. In each case actual series is that of the month in the following year to the last one used for the construction of the model.

3. Results

The last four columns of Table 2 show the indices of the ARMA models identified in each case, and the values of the computed coefficients for each one of them. In six of the sixty series analyzed it was impossible to find an ARMA model that represented them and satisfied the Box–Pierce criterion at a level of significance of 0.1, as recommended by Box and Jenkins (1976).

There are up to ten different ARMA models among those identified, being the most frequent: (1,2) in 37% of the cases; (1,1) in 29.6% and (1,3) and (1,4) in 9.26% of the cases each, with two cases in which the indices of the model, and therefore the number of parameters needed, is very high: (1,7) and (1,8).

As indicated, the study was at first applied to a larger number of stations, for which the same analysis has been carried out, showing a certain tendency in those stations with higher average wind speed (Trinidad, Aralar, Lomanegra, Ujué and Yugo) and consequently of higher interest for the production of energy, to develop models with a higher number of parameters, although we have to point out that it is also in these stations where the amount of available data was higher. In any case the most frequent models in all stations are those with (1,2) and (1,1) orders.

Regarding the forecasts, we include Fig. 4, which displays the evolution of the RMSE and the MABE when the forecast is done 1–10 h in advance, both with the adjusted ARMA model and with the persistence model, and for the month and station indicated. The behaviour observed in this particular case is repeated in the vast majority of the other cases, i.e. the errors obtained with the forecasts done with the ARMA models are always smaller than the ones obtained with the persistence models, as only in four cases, and all of them for predictions 1 h in advance, the persistence model is better than the ARMA model. This result differs from that obtained by Geerts (1984) who came to the conclusion that the RMSE with the persistence model for forecasts 1 or 1 h in advance, were always smaller than those obtained with the ARMA models. In contrast, results are similar to those obtained by Milligan et al. (2003) who also indicate that ARMA models provide significant improvements over persistence models even for short forecast periods.

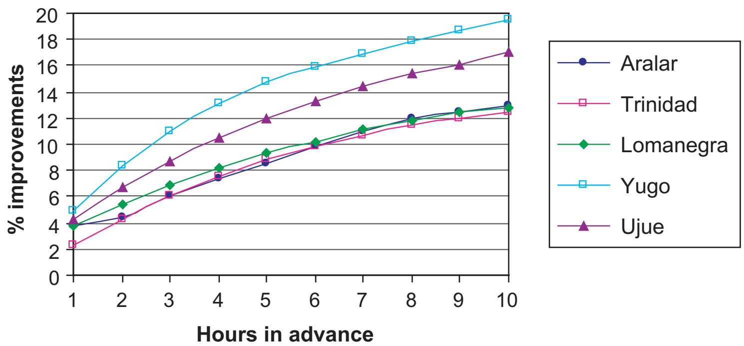


Fig. 6. Percentage of improvement of the RMSE (averaged over all the months) of the forecast when the ARMA model is used, as opposed to the persistence model.

Table 3
Significant values of the RMSE, by station

	Aralar	Trinidad	Lomanegra	Yugo	Ujue
RMSE min m/s	0.97	1.05	1.05	0.90	1.18
Month	August	March	August	October	January
Av(RMSE 1 h) m/s	1.41	1.24	1.19	1.09	1.43
Desv(RMSE 1 h)	0.26	0.19	0.11	0.12	0.24
RMSE max m/s	5.56	3.77	4.00	3.00	3.81
Month	July	November	November	January	February
Av(RMSE 10 h) m/s	3.42	3.02	3.40	2.42	2.89
Desv(RMSE 10 h)	0.98	0.46	0.32	0.28	0.34
Av(RMSE 5 h) m/s	2.83	2.50	2.55	2.06	2.47
DESV(RMSE 5 h)	0.66	0.35	0.25	0.21	0.29

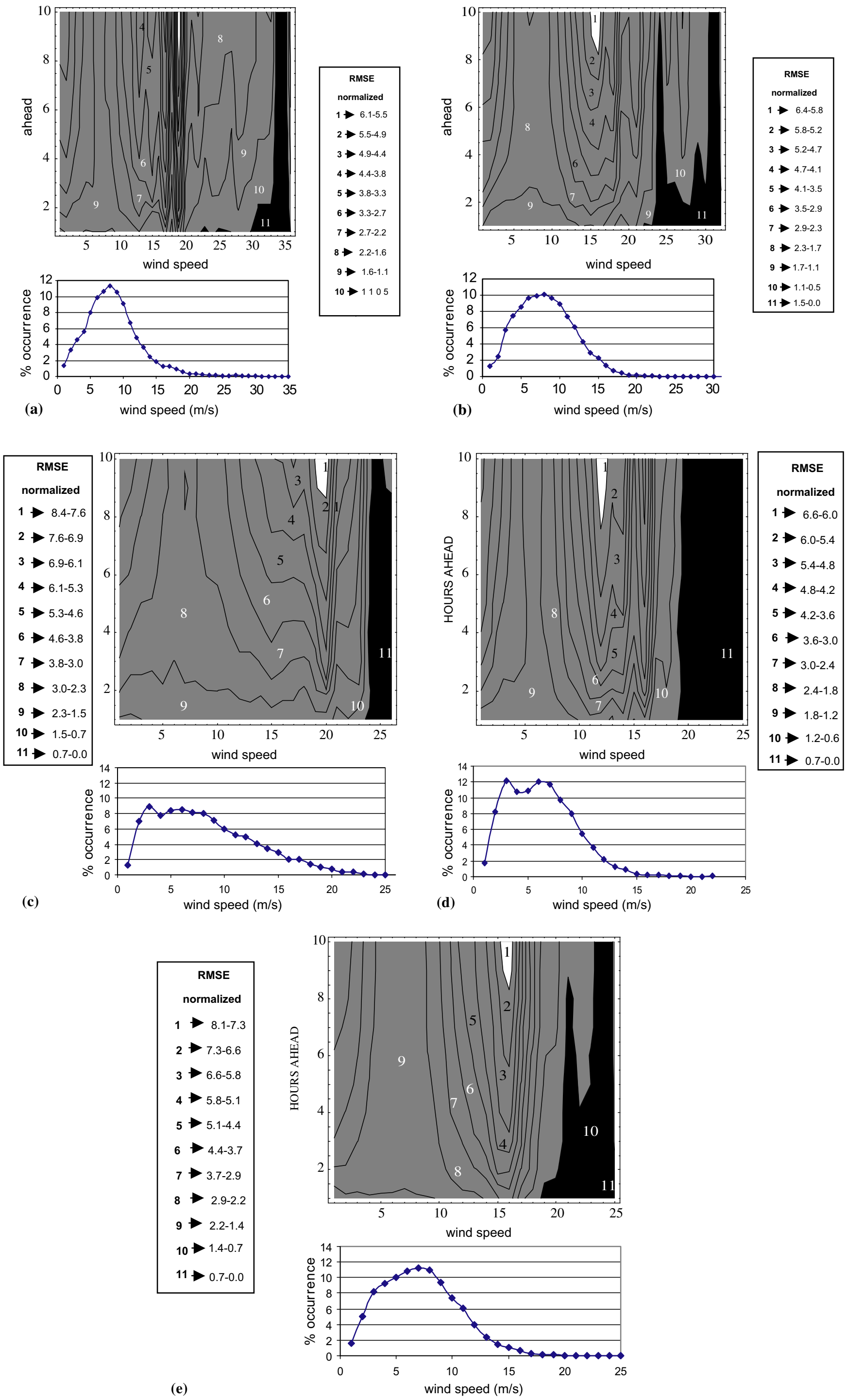


Fig. 7. Plotting of the yearly standardised RMSE values for each wind speed range and for the indicates hours in advance in Aralar (a), Trinidad (b), Lomanegra (c), Yugo (d) and Ujué (e) stations. In the lower graph, the percentage of occurrence of the velocities for each of the ranges is included.

In any case, as we expected, both the RMSE and the MABE increase as the moment of the forecast moves away from the last observed datum available. To that effect, Fig. 5 represents the evolution of the RMSE in the various months studied for the different stations, when the forecast goes from 1 to 10 h in advance, and Fig. 6 shows the improvement in percentage of the average RMSE for each station and advance period of the forecast when the ARMA model is used as opposed to the persistence model. There is an improvement of 2–5% for forecasts 1 h in advance, and 12–20% for 10 h in advance.

In Table 3 we have included, for all the months studied at each station, the minimum value (and the respective month), the average value and the standard deviation of the RMSE for forecasts 1 h in advance; the maximum value (and the respective month), average value and the standard deviation of the RMSE for forecasts 10 h in advance and the average value and the standard deviation of the RMSE for forecasts 5 h in advance. As it can be noticed, the minimum values are in the 1 m/s range, the maximum values in the 3 m/s range, and for forecasts 5 h in advance, in the 2.5 m/s range, so that normally, when doing a forecast more than a few hours in advance, the RMSE are over 1.5 m/s, the limit that Geerts considers acceptable.

From a wind energy generation point of view, the errors made are more important at wind velocities at which the wind turbines are working, so we have evaluated the RMSE at wind speed intervals of 1 m/s, from the minimum to the maximum observed in each actual time series. In order to obtain summarized information of each one of the five stations studied, we have normalized the values obtained each month in the different speed intervals and for each given time in advance with respect to the highest value observed. Finally, we have added up the normalized values of the different months for each speed interval and advance period, resulting in the charts shown in Fig. 7.

4. Conclusions

The process of transformation and standardization of the hourly wind speed time series used provide the necessary conditions to represent them using ARMA models, as can be proved by the fact that in most cases it was possible to identify and validate the appropriate model.

The behaviour of the forecasts in the transformed series is similar to that of the actual velocities, so we can conclude that the use of the hourly mean and standard deviation values used on a monthly basis for the standardization is accurate.

The deterioration of the accuracy of the forecast as the forecasting period increases indicates that if the RMSE is not to exceed the 1.5 m/s threshold, the method applied is only valid for very short-term forecasts.

The use of ARMA models improves significantly the wind speed forecasts as compared to those obtained with persistence models. In only very few cases the errors with the latter are smaller than the ones obtained with ARMA models, and always for forecasts 1 h in advance. In fact, for forecasts 10 h in advance, the errors with ARMA models are between 12% and 20% smaller than with persistence models.

We have not found significant differences in the behaviour of the models for the five locations, either mountainous regions or flat areas, where the data were collected by the weather stations.

The RMSE obtained separately for different wind speed demonstrate differences depending on the wind speed values, showing the following tendencies: Firstly, for wind speeds within a range of 4–11 m/s (that generally correspond to the ascending part of the power curve of the wind turbines, between the cut-in and rated speeds), the errors have relatively low values and are hardly affected by the extent of the forecast. Secondly, the biggest relative errors tend to fall within high wind speeds, higher than the rated speeds of most wind turbines, and for which the power regulation mechanisms of the wind turbines are usually activated. Thirdly, the lowest obtained RMSEs correspond to the highest observed wind speeds. In most cases, wind turbines are stopped for safety reasons at this speed range. This fact, together with the low percentage of occurrence of high wind speeds, makes conclusions in this speed range to be of little significance. In addition, the low value of the normalized RMSE is favoured by this low occurrence which may distort the conclusion.

Acknowledgments

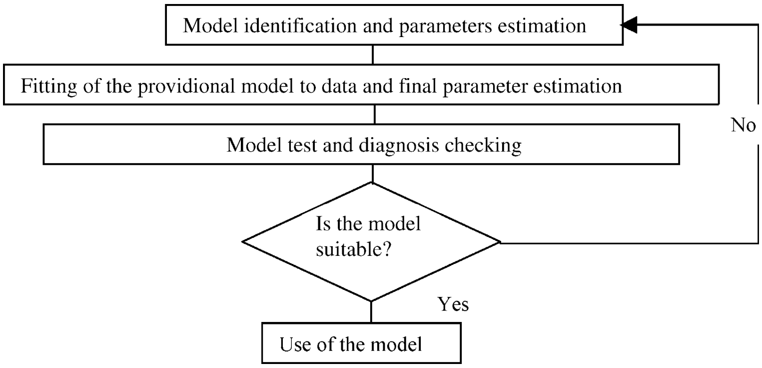
The present study was carried out under a research project supported by the Department of Culture and Education, Government of Navarre, Spain. Authors would like to thank Eloisa Izco for her help with tables and figures.

Appendix A

A.1. General equation of ARMA model

$$(1-\phi_1\cdot B-\phi_2\cdot B^2-\cdots-\phi_p\cdot B^p)\cdot V^*_{n,y} = (1-\theta_1\cdot B-\theta_2\cdot B^2-\cdots-\theta_q\cdot B^q)\cdot a_{n,y}. \tag{A.1}$$

A.2. Algorithm for ARMA models applicability



Appendix B

Mathematical expressions for Durbins relations:

$$\varphi(n,n)=\begin{cases} r_1 & l=1,\\ \frac{r(l)-\sum_{j=1}^{l-1}\varphi_{l-1,j}\cdot r_{l-j}}{1-\sum_{j=1}^{l-1}\varphi_{l-1,j}\cdot r_j} & l=2,3,\ldots,L,\end{cases}$$

$$\varphi_{i,j}=\varphi_{l-1,j}-\varphi_{l,l}\cdot\varphi_{l-1,l-j}\qquad i=1,2,\ldots,L-1. \tag{B.2}$$

References

Balouktsis, A., Tsanakas, D., Vachtsevanos, G., 1986. Stochastic simulation of hourly and daily average wind speed sequences. Wind Engineering 10 (1), 1–11.

Beguin, J.M., Gourieroux, Ch., Monfort, A., 1980. Identification of a mixed autoregressive-moving average process: the corner method. In: Anderson, O.D. (Ed.), Time Series. North-Holland, pp. 423–436.

Box, J.E.P., Jenkins, G.M., 1976. Times Series Analysis, Forecasting and Control. Holden-Day, San Francisco.

Brown, B.G., Katz, R.W., 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.

Chou, K.C., Corotis, R.B., 1981. Simulation of hourly wind speed and array wind power. Solar Energy 26 (3), 199–212.

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

Dubey, S.D., 1967. Normal and Weibull distributions. Naval Research Logistics Quarterly 14, 69–79.

García, A., Torres, J.L., Prieto, E., De Francisco, A., 1998. Fitting wind speed distributions: a case study. Solar Energy 62 (2), 139–144.

Geerts, H.M., 1984. Short range prediction of wind speeds: a system-theoretic approach. In: Proceedings of European Wind Energy Conference. Hamburg, Germany, pp. 594–599.

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

Milligan, M., Schwartz, M., Wan, Y., 2003. Statistical wind power forecasting for US wind farms. NREL/CP-500-35087, November 2003.

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

Sfestos, A., 2000. A comparison of various forecasting techniques applied to mean hourly wind speed time series. Renewable Energy 21, 23–35.

Tsay, R.S., Tiao, G.C., 1984. Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstatitionary ARMA models. Journal of the American Statistical Association 79 (385), 84–96.

Torres, J.L., García, A., Prieto, E., De Francisco, A., 1999. Characterization of wind speed data according to wind direction. Solar Energy 66 (1), 57–64.