Monografías del Semin. Matem. García de Galdeano. **27:** 249–258, (2003).

Physical-Mathematical wind prediction model

José Antonio Escudero Lázaro[1], David Lahoz Arnedo[2]

[1] Dpto. de Informática e Ingeniería de Sistemas. Centro Politécnico Superior (CPS). Universidad de Zaragoza

[2] Dpto. de Métodos Estadísticos. Facultad de Ciencias. Universidad de Zaragoza

joseanto@posta.unizar.es, davidla@posta.unizar.es

Abstract

The aim of this paper is to present an advance of the study carried out to develop a Physical-Mathematical wind prediction model in some zones of the Ebro Valley. This study is divided into several phases. It begins with the capture, filtering and descriptive study of the necessary series of data. Next, some of the variables of interest are fitted to known probability distributions, such as Weibull or Exponential. In the last part, neural networks techniques are used on the series of data. Mainly, two related networks are used. The first one studies "time" prediction, estimating the velocity and the direction of the wind in a place after a lapse of time. The other network studies "spatial" prediction, estimating the same variables in different places at the same instant of time.

Keywords: Neural Networks, Multilayer Perceptrons, SYNOP, Feedforward Networks.

AMS Classification:

1 Introduction

This paper is an advance of the study carried out to develop a Physical-Mathematical model implemented for its computer processing with which, using meteorological data parameters at the mesoscale level, a forescast of the speed and direction of wind is obtained as a result in certain zones of the Ebro Valley.

The guidelines that have been followed until the moment for the accomplishment of the work correspond with the different parts of this paper, and they are the habitual ones in a work of statistical character: taking of data, data processing, estimation, contrast and construction and validation of theoretical models.

In the first part of the paper, meteorological data that are going to be used are described and the results of a descriptive study are shown. Prior to their introduction in the models, the data are filtered, as much by statistical techniques merely as by criteria provided by meteorological experts.

The following section shows some of the techniques used for the fitness of the data series to well-known probability distributions, as well as their results and the validation of these. The study is only for the Weibull and Exponential probability distributions . The first of them is, according to literature on the subject, the one that better adjusts to the distribution of the wind speed [6], whereas the second is necessary for the use of Excesses Over Thresholds techniques. The section ends up showing the results of the techniques used for the adjustments validation.

The last part of the paper describes the neural networks techniques used for the prediction of the wind speed and wind direction. Due to the characteristics of the data series and the aim of the study, it was decided to use Multilayer Perceptron networks. Basically, two combined networks are used. The first one of them acts as a "time expert", predicting the speed and direction of the wind in a place after a lapse of time. This network connects with another one, which acts as a "space expert", estimating the same variables in different places at the same instant of time.

2 Taking of data: types and exploratory analysis

The study has dealt with different meteorological data types. These include METAR, TAF, TEMP, SYNOP data and those that are going to be called ZE data, measured in the zones on which it is desired to make the predictions. In this paper, only the two last types will be used. Although SYNOP data contain information on diverse meteorological variables, those of greater interest for this study are the wind speed and wind direction. The measurements are made by the National Institute of Meteorology in a certain place, the airport of Zaragoza in this case, every six hours beginning at zero hours. The series contains data from the year 1972 to the present time. The series of ZE data begins in 1995 and includes, among other variables, information on wind direction and wind speed with a frequency of ten minutes.

2.1 Exploratory analysis of synop data

In a first contact with the data and after its decoding, its filtering, location of anomalous data, repeated dates, empty fields and missing data were carried out. After eliminating the outliers data (directions greater than 360, negative speeds, speeds superior to a certain value) and the repeated dates, a calculation of the incomplete data was made (approximately a 5%). Not only the analysis of the speed series but also the consultations with expert meteorologists, took to consider out of rank those measurements higher than 25 m/s.

Graphical analyses were made with the objective of discovering as much the predominant wind directions as the wind speeds in those directions. The results were those expected in Zaragoza: two predominant directions that correspond to winds known as "cierzo" (between 270 and 320) and "bochorno" (between 90 and 120). Take a look at *figure 1*.

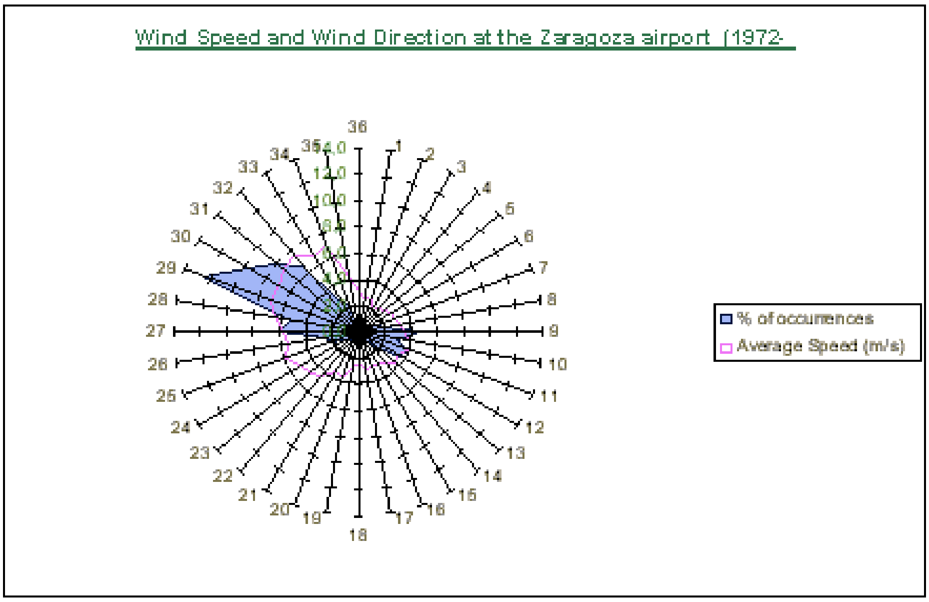


Figure 1: Predominant wind directions at the airport of Zaragoza.

The greater percentage of wind occurrences higher to a certain intensity occurs in the rank of directions corresponding to the *cierzo*.

The series was completed with diverse techniques, from the division in subseries to the filling with the average of the series. The differences between the resulting completed series were not significant for the study, due to the low percentage of missing data. A sample of the difference between the completed series of wind direction in Zaragoza according to several techniques occurs in *table 1*.

In the first column, the statistics of the series without completition is shown, whereas it is completed by means of diverse techniques in the rest columns.

Not to introduce non-real data in the series the study with two series was chosen to be carried out: first, the original one filtered and second, the completed one by means of the average between the four previous data and four later ones.

WIND SPEED IN THE AIRPORT OF ZARAGOZA

		Original	Tendency linear	Interpolation linear	Median of the 4 adjacent ones	Mean of the 4 adjacent ones	Global mean
N	Valid	38324	40279	40279	40279	40279	40279
	Missing	1955	0	0	0	0	0
Mean		4.9854	4.9886	4.9990	4.9949	5.0035	4.9854
Median		4.1000	4.6000	4.1000	4.1000	4.1000	4.6000
Standar deviation		3.7044	3.6139	3.6865	3.6644	3.6584	3.6134
Asymmetry		.734	.749	.731	.732	.728	.752
Curtosis		.130	.286	.130	.156	.159	.290
Minimum		.00	.00	.00	.00	.00	.00
Maximum		21.60	21.60	21.60	21.60	21.60	21.60
Quantile	25	2.1000	2.1000	2.1000	2.1000	2.1000	2.1000
	75	7.2000	7.2000	7.2000	7.2000	7.2000	7.2000

Table 1: Statistics for the SYNOP series of speed (m/s) of the wind at the airport of Zaragoza completed according to different techniques.

2.2 Exploratory analysis of ze data

The filtering of the wind speed and wind direction series in the zones of study is analogous to those made in the previous section. The number of anomalous or absent data is less to one thousand within a series of 231.653 data.

In *table 2* the average speed depending on the average horizontal direction of wind at 40 metres (*cierzo*, *bochorno* or other winds) appears, as well as the distribution of data for each one of them.

	Frequency	Average Speed (m/s)
Cierzo (300-340)	2925	9.27
Bochorno (125-155)	1097	6.61
Other	2119	5.17

*Table 2:*Relation between average speed and wind direction in ZE data.

With the aim to relate ZE data with the data in the SYNOP reports of the airport of Zaragoza, the data that correspond to the hours of these reports (0, 6, 12 and 18 hours) have been extracted.

3 Adjustments to models

Two types of adjustments have been achieved, by means of EOT techniques ("Excesses over Thresholds") and by means of adjustments to well-known probability distributions.

3.1 Adjustment to a eot model

The objective is to take advantage of the Excesses Over Thresholds statistical techniques(EOT) in order to create a model of prediction of the wind speed in the airport of Zaragoza. The study has been made on the complete series and on series obtained by dividing the former one in periods of more homogenous behaviour. The EOT technique purpose is to verify if the times between two consecutive increasing crossings over a prefixed threshold are independent and equally distributed exponential accomplishments [7]. If we are able to prove this fact for a certain threshold using the wind speed data in the airport of Zaragoza, it is possible to suppose that the number of times by time unit that the speed is over that threshold is a Poisson or Generalised Pareto process.

Several tests dividing the SYNOP series of the wind speed in Zaragoza in different periods of time, such as seasons, months and so on have been carried out. The results that this paper shows correspond to the fitness to exponential distribution with the series between January the 20" and April the 20". The objective is to find the threshold for which the distribution of this series fit to an exponential distribution.

There are three techniques to verify the goodness of the adjustment : histogram, Q-Q plot and Kolmogorov-Smirnov Test. *Figures 2* and *3* show different histograms and Q-Q plot on thresholds of 4 and 9 m/s of the crossing data in the three months mentioned above.

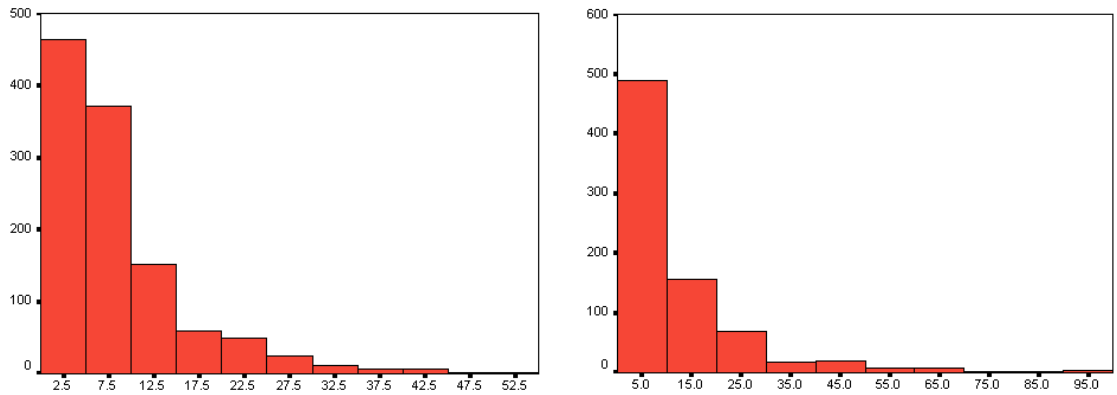


Figure 2: Distribution of the data of time over threshold 4 m/s (left) and 9 m/s (right) in the SYNOP series (periods from January the 20" to April the 20"). Each unit of the abscissa axis corresponds to six hours.

The last test is a non-parametric test: the Kolmogorov-Smirnov Test. Data in table 2 show the level of significance with which it is possible to accept that the distribution of excesses over a threshold adjusts to a exponential distribution according to this test.

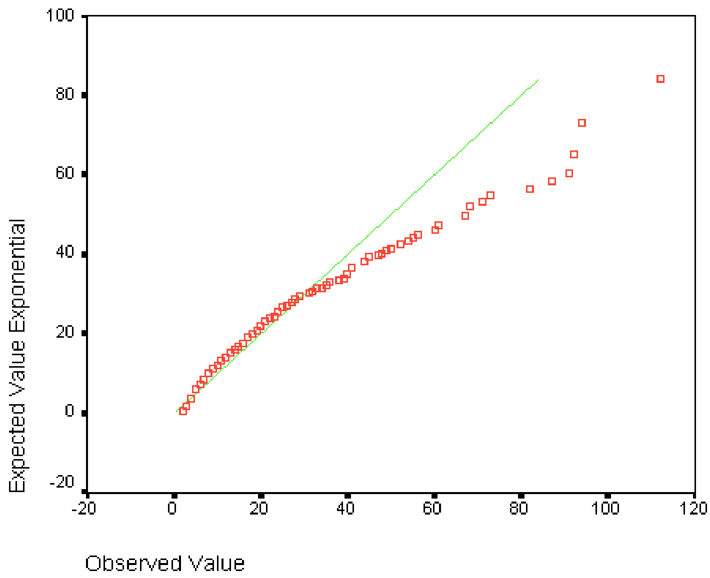


Figure 3: Q-Q plot for threshold 9 m/s in the SYNOP series (periods from January the 20" to April the 20"). Each unit corresponds to 6 hours.

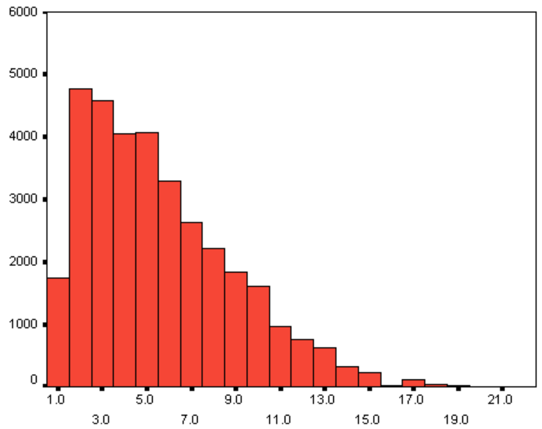
Thresholds on m/s.	Level of significance
3	0,0003
4	0,0002
5	0,0002
6	0,0001
7	0,0002
9	0,0002
11	0,0003

Table 2: Level of significance of the Kolmogorov-Smirnov Test for the series of time over different thresholds in the SYNOP series(periods from January the 20" to April the 20").

3.2 Adjustment to the weibull distribution.

In this part of the study the wind speed in the airport of Zaragoza (SYNOP data) is attempted to be fitted to well-known distributions, as Weibull, Gamma or Exponential. The best results are obtained for the Weibull distribution, which is the reason why this will be the only one to appear in this paper. As the Weibull distribution only takes positive values, the adjustment has been made only on the cases in which the speed is greater than zero. The selection of this distribution to make the adjustment is due to the similarity between the histogram of the wind speed variable in the airport of Zaragoza and the histogram of a Weibull distribution.

The tests for the goodness of the adjustment are the same ones that have been used to fit the distribution of excesses over thresholds to the exponential distribution: histogram, Q-Q plot and Kolmogorov-Smirnov Test. The parameters of the Weibull distribution corresponding to the series are 6.253 and 1.823 for scale and shape respectively. The results of these analyses are in *figure 4* and *table 3*.



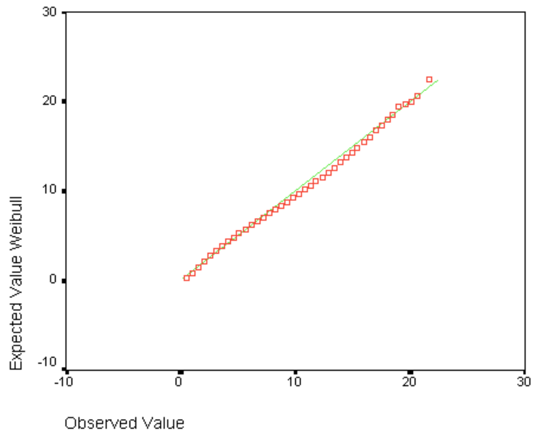


Figure 4: Histogram of the positive speed series(m/s) in the airport of Zaragoza and Q-Q plot of this same series with adjustment to a Weibull distribution of scale 6.253 and shape 1,82.

	Size of sample	Value of K-S statistic	p- value for 0.01 and 0.2
All data	35859	0.08	0.008- 0.005

Table 3: Level significance of the Kolmogorov-Smirnov Test for the series of positive speed in the airport of Zaragoza. Distribution of contrast: Weibull of scale 6.253 and shape 1.823.

4 Wind prediction with neural networks

The idea is to look for a method that allows to predict the speed and the direction of the wind in certain zones of the Ebro Valley from SYNOP data of Zaragoza with the minimum possible error. The first methods used are the linear and curvilinear regression, to later go on to use methods based on neural networks wich adapt better to the behavior of the data and provide better results than the previous ones.

A neuronal network associates, by means of functions and weights, some variables (called input) with others (called output). Neural network techniques are used not only to predict the output variables but also to recognize patterns or groups within the data.

Using the first of this ideas, a neural model to predict the wind speed and the wind direction in a zone, from the same variables in the airport of Zaragoza, have been constructed. There are different network models according to the technique that is used. In this paper the Multilayer Perceptron is used (MLP).

In order to make the prediction, two combined networks are used. The first one acts as a "time expert", predicting the wind speed and the wind direction in the airport of Zaragoza for a moment T from the measurements of speed and direction in the same place at moments T -6, T -12, T -18 and T -24 hours. This network connects with another one, that acts as a "spatial expert", which considers the speed and the direction of the wind at moment T in a ZE zone from the prediction made by the time expert for the same moment T.

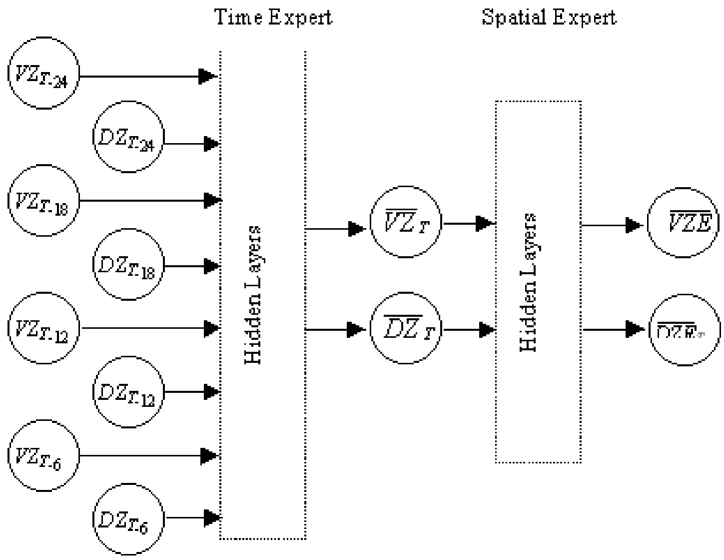


Figure 5: General scheme of the neural networks used in the study. The meaning of the symbols is the following one: VZ_t, DZ_t are the wind direction and wind speed data in the airport of Zaragoza at moment t; $\overline{VZ_t}$ and $\overline{DZ_t}$ are the predictions of wind speed and wind direction in the airport of Zaragoza at moment t; $\overline{DZE_t}$ and $\overline{VZE_t}$ are the predictions of speed and direction of the wind in ZE zones at moment t.

The first neural network corresponds to the "spatial expert" due to its simplicity.

4.1 Spatial expert

In a neural network, it is necessary to fix the existing relation between the input variables and the hidden ones, and between these and the output variables. In this study, the input variables are wind direction and wind speed in Zaragoza, from which the prediction is going to be carried out. The output variables are the speed and direction of the wind in the zones of study.

Having fixed the input and output variables, the next step is to determine the number of hidden nodes that are introduced in the model, as well as the number of layers. Most of the theoretical texts on neural networks advises to take a single hidden layer [1-2-3]. The difference between taking one or several is not significant at the time of predicting. According to what has been verified, when the number of hidden layers increases, the difficulty of the model without improving the prediction increases as well . Once the structure of the network has been decided, the initial weights and the transference functions, that relate the different nodes and layers, must be fixed. The more usual functions of transference between the input layer and the hidden layer are the hyperbolic tangent and the logistic functions[4], with not significant differences between both for models with structure 2-j-2 [1-2]. In this study the hyperbolic tangent function has been selected. Between the last hidden layer and the output layer the transference function is a linear function, taking into account that nonlinear behaviour is in the previous layers. As a rule of learning for the calculation of the weights the conjugate gradient technique has been used [1-2-3]. In order to avoid scale problems the input and output data have been standardized and the initial weights have been obtained from random uniform distribution between -0,1 and 0,1[5]. The neural network scheme for the spatial expert used in the study is show in *figure 6.*

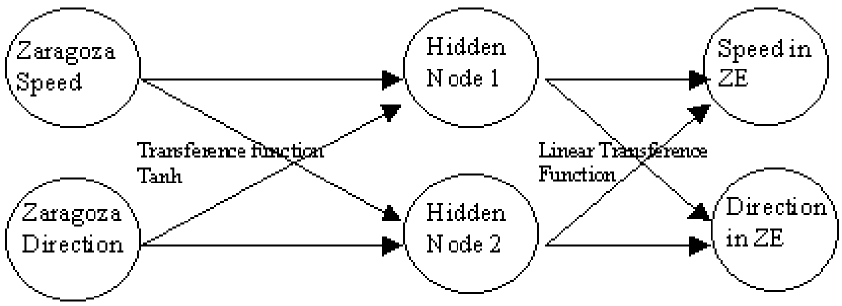


Figure 6: Structure of the neural network used in the study as "spatial expert".

As it can be verified in section 2.1, the behaviour of the wind in the airport of Zaragoza is different according to the direction. For this reason, the study was completed by constructing four different networks, each one with the quadrant data as input data.

4.2 Time expert

The process followed for the construction of this neural network is analogous to the one of the previous section. The difference between both is in the structure, 8-3-2 in this case. A MLP model with conjugate gradient technique as learning algorithm, standardization of the input and the output and initialisation of weights with the uniform distribution between -0.1 and 0.1 have been chosen. The transference or activation function between the input layer and the hidden one is the hyperbolic tangent function, and between this latter one and the output layer the linear function is used.

References

[1] Terrence L. Fine. *Feedforward Neural Networks Methodology*. Springer 1999

[2] B.D. Ripley. *Pattern Recognition and Neural Networks. Cambridge.* 1996.

[3] Teuvo Kohonen. *Self-Organizing Maps.* Springer 1994.

[4] W. Duch, N. Jankowski. *Survey of Neural Transfer Function.* Neural Computing Surveys 2. 1999.

[5] G. Thimm, E. Fiesler. *H*igh order and Multilayer Perceptron Initialization. IEEE Transactions on Neural Networks. 1996.

[6] H. Bergström and A. Smedman. *Wind climatology for a well-exposed site in the Baltic Sea.* Wind Engineering Vol. 23 N 3. 1999.

[7] A. C. Cebrián. *Análisis, modelización y predicción de episodios de sequía.* Memoria doctoral. 2000.