Monthly Forecasts of the Average Wind Speed in Portugal

J. Rio[1], T. Simões[2], A. Estanqueiro[2]
[1] IM – Instituto de Meteorologia, http://web.meteo.pt
[2] INETI – Instituto Nacional de Engenharia, Tecnologia e Inovação, http://www.ineti.pt
joao.rio@meteo.pt, teresa.simoes@ineti.pt, ana.estanqueiro@ineti.pt

Abstract:

Forecasts of the monthly anomaly of the average wind speed can provide useful information for energy management and production. Unlike methods which use state-of-the-art highly complex dynamical or statistical models, this study presents a methodology based on analogues. This simple method identifies similar weather patterns using past observations to produce a forecast and relies on the fact that the atmosphere has a limited number of preferred weather patterns. The forecast is provided as the most likely category and its skill is assessed against persistence. Results are shown only in an area limited to Portugal and the nearby Atlantic Ocean.

Keywords: **Monthly forecast, average wind speed, skill.**

1 Introduction

Power System operation requires wind speed forecasts of mainly two types: short range (24-48 hours) and long range (one month and beyond). The first type of forecasts can provide useful information on the grid integration of energy from wind parks, as well as extreme weather affecting wind park production. The second type of forecasts can provide useful information to increase the efficiency of system management, as it may be able to identify periods where the total energy output of wind parks is expected to be higher/smaller than average.

The atmosphere is a caotic system, that is, no matter how small the amplitude of initial errors, they eventually increase and lead to a decorrelation between forecast and observation after some time. The errors arise from the impossibility of knowing the exact state of the atmosphere and the parametrization of physical processes and mathematical issues. For this reason, the atmosphere has a limited predictability, which implies that forecasts beyond ten days are probabilistic in nature. Also as a consequence, long range forecasts do not and can not depict individual weather, but provide the expected average conditions for a given period of time.

Presently, long range forecasting is one of the most important research fields. It is based on the existence of atmospheric large scale phenomena which exhibit slow variability and/or recurrence and the persistence of the boundary conditions. This kind of forecasts is mainly produced by state-of-the-art dynamical models, which require an extremely high computer power, or by statistical modelling. The skill of long range forecasts is much lower than in the short and medium range. Generally, skill is higher in the tropics than in the mid-latitudes, being particularly limited in Europe [1].

In this study a forecast method based on the analogues technique is developed. The objective of this technique is to provide a forecast based on past observations and relies on the fact that the atmosphere has a limited number of preferred states, the so called weather regimes. The two main advantages of this methodology are its simplicity and low computational cost. On the other hand, the major drawbacks are the need for a very long data set and the difficulty of finding true analogues. Additionally, even if one can find a good analogue there is no warranty that it will lead to a good forecast because of the irregular behaviour of the atmosphere.

The forecasts provide the probability of the average wind speed being 'above normal', 'near normal' or 'below normal'. From this set of probabilities, a categorical forecast is produced. The forecast skill is assessed and compared against climate and persistence forecasts in the area of Portugal (Continental, Azores and Madeira).

2 Methodology

2.1 Data

The atmospheric data used was NCEP/DOE Reanalysis II [2] was obtained from NOMADS [3]. The data covers the period January 1979 to December 2005, providing forecasted wind speeds four times a day (00, 06, 12 and 18 UTC). The data sets used in this study are the geopotential height of the 1000 hPa (hereafter designated by hgt1000), the 10 m and the 925 hPa wind speeds. These two variables can be considered representative of the wind speed at low level and medium level sites (altitudes of 750-1000 m), respectively. The wind speed at 10 m is provided in a Gaussian grid, with a resolution of approximately 1.875º x 1.903º. The hgt1000 and the 925 hPa data sets are provided in a 2.5º x 2.5º lat-long grid. The daily average was calculated for all the data, to obtain a single value per day.

For each day of the year, the climatology of the two wind speed data sets is calculated as a thirty day average, from which the percentiles 33% and 66% are determined.

The Smith and Reynolds Extended Reconstructed Sea Surface Temperature [4] was used, covering as well the period January 1979 to December 2005. This data is provided as the monthly average in a 2º x 2º lat-long grid and is intended for boundary condition information.

A Principal Component Analysis was then performed on both the hgt10000 and SST data sets to identify and retain only the first modes of variability of the fields. The analysis was done in the domain [(20ºN, 100ºW); (80ºN, 60ºE)], with no distinction between seasons.

2.2 Analogue Model

The predictors of this method are based on the daily principal components of the hgt1000 field (hgt-pc, hereafter) which represents the atmospheric circulation. The similarity between the weather pattern at the given initial date and each day of the history record is based on the following four parameters.

The first parameter represents the initial day weather circulation and is done by assign a vector with the first N principal components of the hgt1000 field. To allow the model to have some knowledge of the recent past evolution of the atmosphere, two additional predictors are used from the hgt-pc data set: the mean conditions and the trend of each one of the pre-defined M pc. These quantities are calculated over a period of L days between t-L and t-1.

The information after the boundary conditions arise from the use of the SST data set. The procedures with this predictor are identical to the ones taken when handling the hgt variable. In this case, as the SST data is provided as a monthly value, a linear interpolation is made to account for the specific day of the initial date.

For a given initial date, the similarity with each day of the historical data is determined objectively as the cosine of the angle formed by two vectors, as described by equation 1.

$$S(t,t') = \cos\left(\frac{\sum_{i=1}^{N} \alpha_i(t)\beta_i(t')}{\|\alpha_i(t)\|.\|\beta_i(t')\|} \right) \qquad (1)$$

where S stands for the similarity between the vectors α and β, given by the first N pc, respectively at time t and t'. This procedure is done for all the four parameters characterising each day. Finally, the difference between the initial date t and each potential analogue at time t' is done by a linear combination of the values $S(t,t')$ calculated previously for each parameter, as described in equation 2.

$$\Delta(t,t') = \sum_{i=1}^{4} \alpha_i S_i(t,t') \qquad (2)$$

After sorting the quantities Δ, a list of the most similar days is obtained, thus providing the analogue days which will provide the forecast. Several restrictions are used to avoid the selection of analogue days to close from each other and also from the given initial date.

Starting from the day of each one of the selected analogues, the average of the variable to be forecasted is calculated over a thirty day period. The value obtained is then compared with the previously calculated 33% and 66% percentiles, allowing the classification into one of the three following classes: 'below normal', 'near normal' and 'above normal'. This procedure is done for each one of the pre-selected total number of analogues to use and thus allows to calculate the probability of occurrence of each of the classes. A categorical forecast is then made by electing the class with the highest probability. Finally, the forecasts are done in a complementary period regarding the one used for the climatology and percentile calculation.

3 Skill Assessment

3.1 *Gerrity Skill Score*

Equitable skill scores are often used to evaluate multi-category forecasts. A recommended skill score for a three by three contingency table which has many of the desirable properties and is easy to compute is the Gerrity Skill Score [5,6].

Table 1: Example of a three by three contingency table.

		Observations		
		Below normal	Near normal	Above normal
Forecasts	Below normal	n_{11}	n_{12}	n_{13}
	Near normal	n_{21}	n_{22}	n_{23}
	Above normal	n_{31}	n_{32}	n_{33}

The relative sample frequencies p_{ij} are defined as the ratios of the cell counts to the total number of forecast/observation pairs N:

$$p_{ij} = {}^{n_{ij}}\!/_{N} \qquad (3)$$

The sample probability distributions of forecasts and observations become, respectively:

$$p(f_i) = \sum_{i=1}^{3} p_{ij} = \hat{p}_i ; i = 1,3 \qquad (4)$$

$$p(x_i) = \sum_{j=1}^{3} p_{ji} = p_i ; i = 1,3 \tag{5}$$

The definition of the score uses a scoring matrix $s_{ij}\,(i = 1,...,3)$, which is a tabulation of the reward or penalty forecast/ observation outcome represented by the contingency table will be accorded:

$$GSS = \sum_{i=1}^{3}\sum_{i=1}^{3} p_{ij} s_{ij} \tag{7}$$

The scoring matrix is given by

$$s_{ii} = \frac{1}{2}\left(\sum_{r=1}^{i-1} a_r^{-1} + \sum_{r=i}^{2} a_r\right) \tag{8}$$

$$s_{ij} = \frac{1}{2}\left[\sum_{r=1}^{i-1} a_r^{-1} - (j-i) + \sum_{r=j}^{2} a_r\right] \tag{9}$$

$$1 \le i < 3; i < j \le 3$$

where

$$a_i = \frac{1 - \sum_{r=1}^{i} p_r}{\sum_{r=1}^{i} p_r} \tag{10}$$

Note that the Gerrity Skill Score (GSS) is computed using the sample probabilities, instead of those on which the original categorizations were based (*i.e.* 0.33, 0.33, 0.33).

The GSS is a recommended score because the ease of construction ensures its consistency from categorization to categorization and with underlying linear correlations. The score is likewise equitable, does not depend on the forecast distribution, does not reward conservatism, utilizes off diagonal information in the contingency table and penalizes larger errors more. Skill scores of 1 indicate a perfect forecast, whereas a skill lower or equal to 0 stands for a forecast no better than climatology.

3.2 Forecasting skill

The skill of the monthly forecasts is made in a domain that covers Portugal (Continental, Azores and Madeira) in each of the four seasons. The initial dates of the forecasts are the 1st and the 16th of each month, therefore there is an overlap of two weeks. The number of forecasts is 158 in winter (December, January and February), 161 in fall (September, October, November) and 162 in summer (June, July, August) and spring (March, April, May).

The forecast variables are the 1000 hPa geopotential height and the wind speed at 10 m and 925 hPa. A sensitivity study with the hgt1000 variable was made to determine the number of several parameters (*e.g.* number of analogues, the linear combination coefficients) that maximize the forecasting skill. Once determined, the forecast of the 10 m and the 925 hPa wind speed were made.

The forecast skill is compared with persistence which is, along with the climatology, the two most basic forms of forecast. For a forecast tool to be valuable it must have a skill higher to both persistence and climatology forecasts.

Figures 1 to 4 show the GSS of the average monthly forecasts of the 1000 hPa geopotential height for each of the seasons.

The skill of 1000 hPa geopotential height is highest in winter, with scores ranging from 0,2 to 0,4. In summer, the scores are lower with large areas exhibiting a skill around 0,2. In the two transition seasons, i.e. spring and fall, the skill is generally lower, with large areas exhibiting marginal or no skill at all.

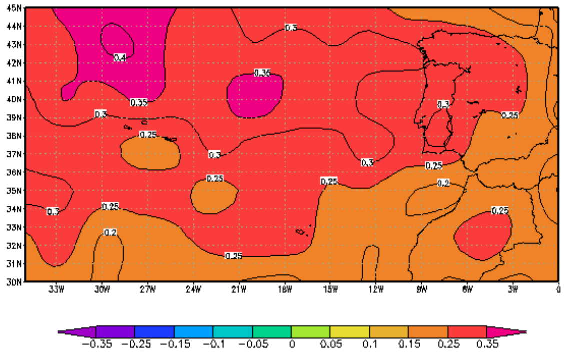

Figure 1: Skill score of the monthly average z1000 hPa forecast in winter.

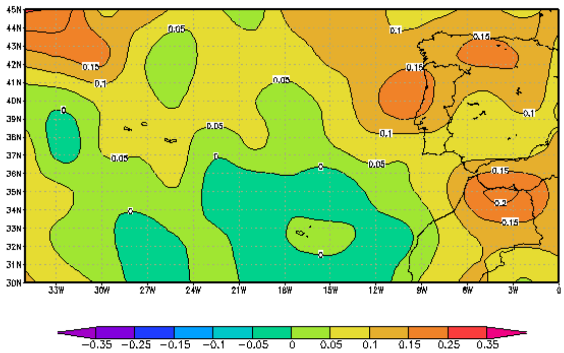

Figure 2: Skill score of the monthly average z1000 hPa forecast in spring.

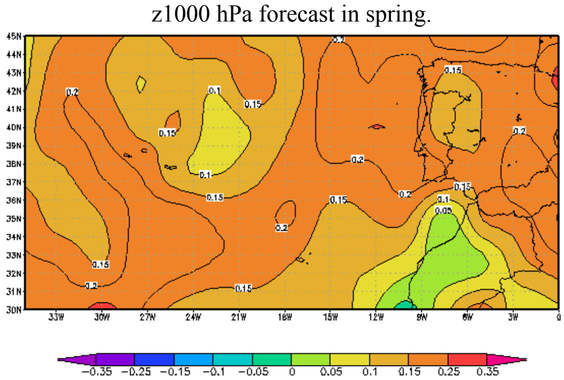

Figure 3: Skill score of the monthly average z1000 hPa forecast in summer.

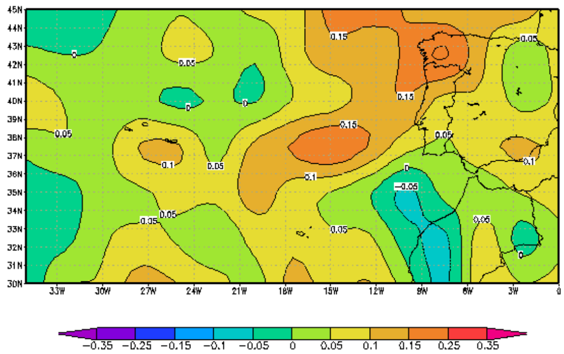

Figure 4: Skill score of the monthly average z1000 hPa forecast in fall.

Figures 5 to 8 present the difference of skill between the model and a persistence forecast, which is a particularly tough test for extended-range forecasting. In the figures below, positive areas indicate that the skill of the model presented is higher than persistence. The skill of the model is generally higher than persistence in winter and summer. Spring and fall are found to be the most difficult seasons as the areas with negative values are quite extensive.

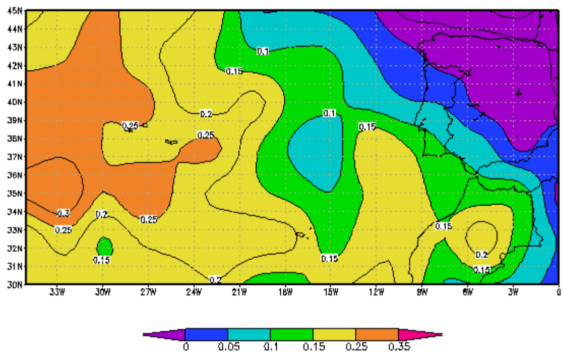

Figure 5: Difference between the skill of persistence and model forecast of the monthly average z1000 hPa in winter.

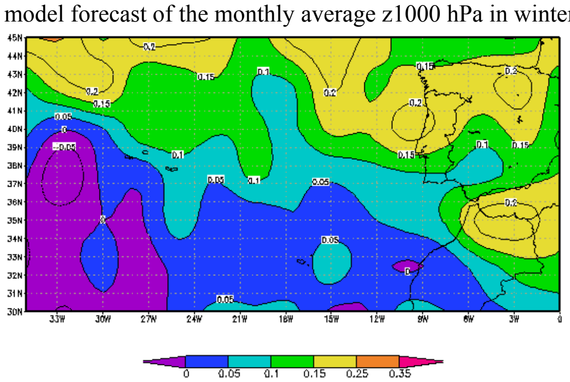

Figure 6: Difference between the skill of persistence and model forecast of the monthly average z1000 hPa in spring.

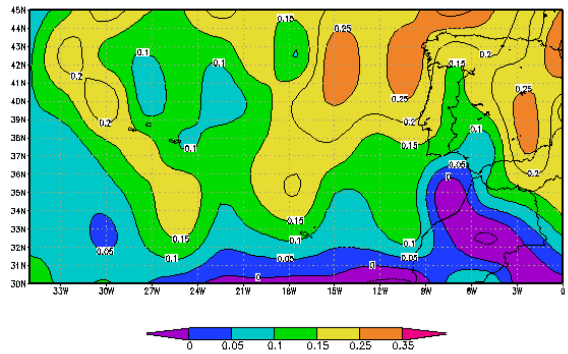

Figure 7: Difference between the skill of persistence and model forecast of the monthly average z1000 hPa in summer.

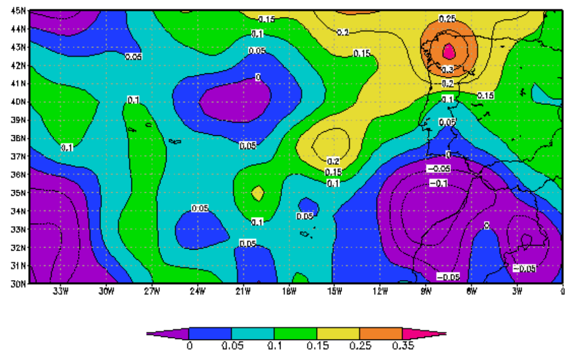

Figure 8: Difference between the skill of persistence and model forecast of the monthly average z1000 hPa in fall.

Figures 9 to 12 present the GSS for the 925 monthly average wind speed. The patterns in the GSS are identical to the ones found in the plots of the 1000 hPa geopotential height, but skill is lower. The skill scores are highest in winter, with the remaining seasons showing no marked differences between each other.
When comparing the model and persistence skill, it becomes clear that apart for some restricted areas, persistence forecasts are as good or even better than the presented model.

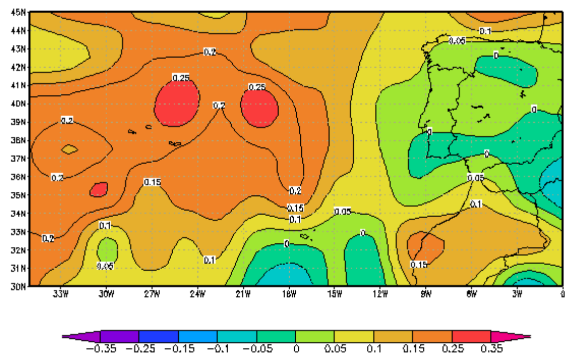

Figure 9: Skill score of the monthly average 925 hPa wind speed forecast in winter.

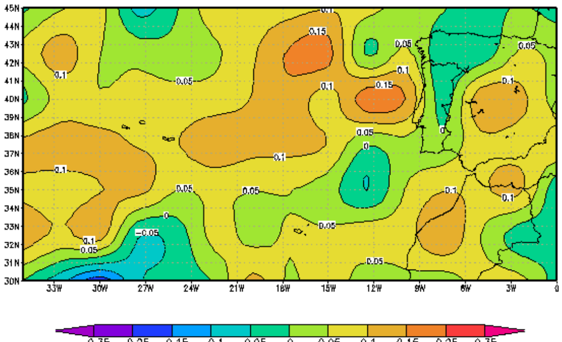

Figure 10: Skill score of the monthly average 925 hPa wind speed forecast in spring.

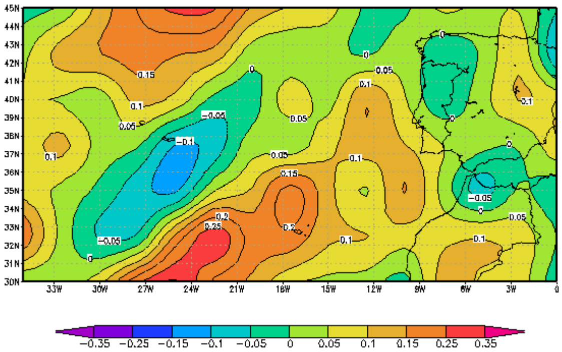

Figure 11: Skill score of the monthly average 925 hPa wind speed forecast in summer.

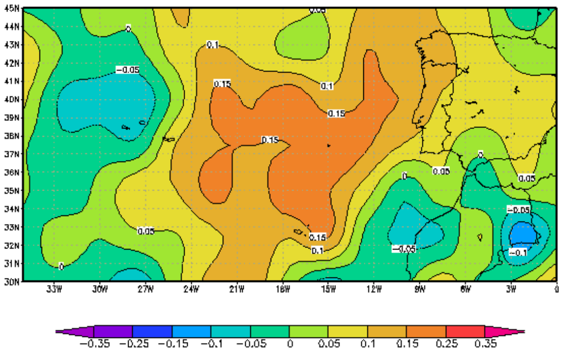

Figure 12: Skill score of the 925 hPa wind speed forecast in fall.

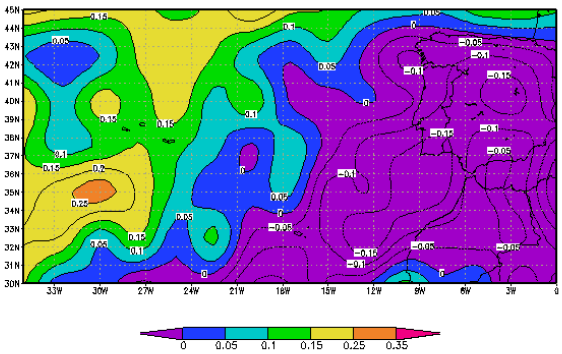

Figure 13: Difference between the skill of persistence and model forecast of the 925 hPa wind speed in winter.

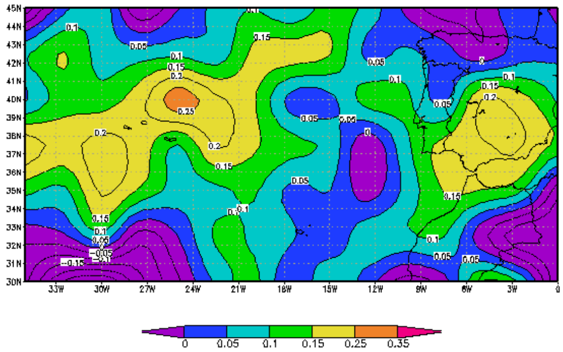

Figure 14: Difference between the skill of persistence and model forecast of the 925 hPa wind speed in spring.

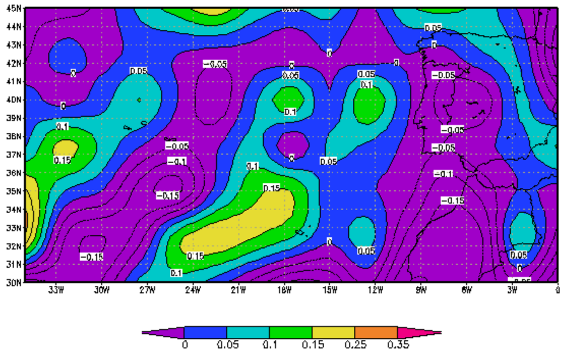

Figure 15: Difference between the skill of persistence and model forecast of the 925 hPa wind speed in summer.

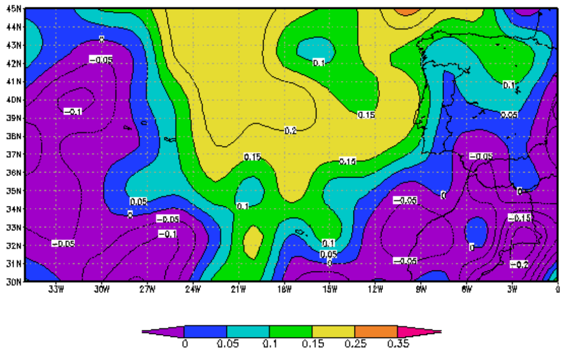

Figure 16: Difference between the skill of persistence and model forecast of the 925 hPa wind speed in fall.

Tables 1 to 3 present the area average of the GSS for the monthly forecasts of the 1000 hPa geopotential height, 925 hPa and 10 m wind speeds, respectively. Two areas are shown: one that is limited to Continental Portugal and the one presented in the previous figures, which is named "Portugal and nearby Atlantic".
The calculated scores show that both the 925 hPa and the 10 m wind speed forecasts made with the model have marginal skill, regardless of the season. However, predictability strongly depends on geographical location and therefore some areas have some useful skill.
For the 1000 hPa geopotential height variable, the model forecasts have a higher skill than persistence, regardless of the area or season. Even though in

spring and fall skill is very limited, the scores for winter and summer are appreciable.

Table 1: Gerrity skill score of the monthly forecast of the 1000 hPa geopotential height.

	Portugal (Continental)		**Portugal and nearby Atlantic**	
	Persistence	Model	Persistence	Model
Winter	0.23	0.29	0.13	0.26
Spring	-0.03	0.12	-0.03	0.06
Summer	-0.03	0.19	0.03	0.16
Fall	-0.01	0.12	-0.01	0.07

Table 2: Gerrity skill score of the monthly forecast of the 925 hPa wind speed.

	Portugal (Continental)		**Portugal and nearby Atlantic**	
	Persistence	Model	Persistence	Model
Winter	0.13	0.02	0.09	0.10
Spring	-0.05	0.04	-0.01	0.04
Summer	0.05	0.03	0.07	0.05
Fall	0.00	0.10	0.00	0.04

Table 3: Gerrity skill score of the monthly forecast of the 10 m wind speed.

	Portugal (Continental)		**Portugal and nearby Atlantic**	
	Persistence	Model	Persistence	Model
Winter	0.12	0.01	0.07	0.08
Spring	-0.03	0.06	-0.02	0.04
Summer	0.10	0.05	0.07	0.06
Fall	-0.02	0.09	0.02	0.07

4 Concluding Remarks

Forecasting beyond the medium-range is the one the main research areas at present and is a particularly difficult task. Such forecasts are usually done with highly complex general circulation models which require an extremely high computer power.

In this study a simple and low cost model is presented to forecast the 1000 hPa geopotential height, the 10 m and the 925 hPa monthly wind speeds. The model is based on an analogue approach, that is, it looks on the historical records for similar patterns to supply a probabilistic forecast. From this set of probabilities a categorical forecast is made.

The skill of the model is assessed with an equitable skill score and is compared against persistence, which is a tougher test for extended-range forecasting than climatology.

The score of the forecasts suggests that the model exhibits some useful skill in forecasting the atmospheric circulation at the 1000 hPa. Skill scores are higher in winter and summer than in the transition seasons. In large areas the model forecast beats persistence.

The 10 m and the 925 hPa average wind speeds are far more difficult, as skill is marginal. Also, apart for some restricted areas in the domain considered, the model forecasts do not beat persistence, even though it is generally superior to climatology.

The skill of these forecasts is in agreement with published results and reinforces the fact that Europe is a particularly difficult area in monthly forecasting.

References

[1].Vitart, F., “Monthly Forecasting at ECMWF”, *Bul. of the Atmos. Met. Soc.*, Dec 2004, 2761-2779.

[2].M. Kanamitsu, W. Ebisuzaki, J. Woollen, S-K Yang, J.J. Hnilo, M. Fiorino, and G. L. Potter. “NCEP-DEO AMIP-II Reanalysis (R-2)”, *Bul. of the Atmos. Met. Soc.*, Nov 2002, 1631-1643.

[3].Rutledge, G.K., J. Alpert, R. J. Stouffer and B. Lawrence: The NOAA Operational Archive and Distribution System (NOMADS), 2002. Proceedings of the Tenth ECMWF Workshop on the Use of High Performance Computing in Meteorology - Realizing TeraComputing. Ed. W. Zwieflhofer and N. Kreitz, World Scientific, pp 106-129, 2003, Reading, UK.

[4].Smith, T. M. and R.W. Reynolds, 2003: “Extended Reconstruction of Global Sea Surface Temperatures Based on COADS Data (1854-1997)”. *J. Climate*, **16**, 1495-1510.

[5].“Standardised Verification System for Long-Range Forecasts”; New Attachment II-9 to the Manual on the GDPS (WMO-No. 485), Volume I; Version 3.0, 2002.

[6].Gerrity, J.P., 1992: “A note on Gandin and Murphy’s equitable skill score”. *Mon. Wea. Rev.*, **120**, 2709-2712.