NCEP/NCAR Reanalysis II for long-term wind speed variability characterization

Rosa Pilão[1], Paulo Pinto[2], Rui Pereira[2,] Ricardo Guedes[2]

[1] ISEP, Instituto Politécnico do Porto
Rua Dr. António Bernardino de Almeida 341, 4200-072 Porto (Portugal)
Phone: +351 22 8340500, Fax: +351 22 9537352, e-mail: rmp@isep.ipp.pt

[2] MEGAJOULE
Rua Eng. Frederico Ulrich, 2650 4470-605 Moreira da Maia (Portugal)
Phone: +351 22 0915480, Fax: +351 22 9488166, e-mail: paulo.pinto@megajoule.pt
ricardo.guedes@megajoule.pt
rui.pereira@megajoule.pt

Abstract

Direct comparison of the local measured data with reanalysis datasets in order to assess the variability of wind speed is often used by the wind industry. A particular concern associated with those analyses is the mismatch between the local masts location and grid points where the reanalysis data are available. This work aims to be a contribution to solve this matter.

Different solutions as to how the reanalysis datasets, taken for the grid points, can be used in order to better represent the variability of the wind speed, at measurement locations, were tested. The idea was to understand how sensitive is the correlation between onsite wind data, and reanalysis data, to different grid point definitions and to different interpolation techniques.

On a second stage, the long-term trends of NCEP/NCAR R2 datasets were evaluated.

In total, 36 wind measurement campaigns located in three different countries and 32 reanalysis datasets were used in this study. The masts selected for this investigation are located in Portugal (15), Poland (11) and Romania (10). These meteorological stations were chosen due to the completeness and longevity of the obtained set of records.

The majority of the tested sites show a reasonable correlation between mean monthly wind speeds, regardless of the interpolation scheme used. Results are also not very different between regions. Sites in Portugal show slightly worst correlations but this could easily be explained with the fact that the terrain in most of those sites is more complex than that from the other countries.

Besides space continuity, the use of NCEP/NCAR reanalysis data for long term corrections of short term measurements demands a high standard of consistency through time. The time consistency was analysed by means of the slope (*a*) of a linear regression ($y=ax+b$) fitted to all NCEP/NCAR 20 year time series. No significant upward or downward trend was found in any of the used data sets, indicating that the reanalysis data series are consistent through time.

Keywords: reanalysis data, onsite measurements, interpolation.

1. Introduction

The wind resource is the fundamental input to any consideration of a wind farm outlook. Particularly in what concerns energy yield estimates the accurate characterization of wind speed variability over the wind farm lifetime is of critical importance [1, 2]. Different approaches exist but all of them have in common the need for onsite measurements and its extrapolation to long term by adequate means.

The commonly used practices, for the long-term extrapolation, are mainly of three kinds: correlation with a nearby reference station with long-term wind data, correlation with virtual long-term data series obtained by mesoscale simulation or direct comparison with reanalysis datasets.

Although, perhaps the least precise, direct comparison of the locally measured data with reanalysis datasets in order to assess the variability of wind speed is often used by the wind industry.

A particular concern associated with those analyses is the mismatch between the local masts location and grid points where the reanalysis data are available [3]. This work aims to be a contribution to solve this matter.

Different solutions as to how the reanalysis datasets taken for the grid points can be used in order to better represent the variability of the wind speed at measurement locations were tested. The idea was to understand how sensitive the correlation between onsite wind data and reanalysis data is to different grid point definitions and to different interpolation techniques.

On a second stage, the long-term trends of NCEP/NCAR R2 datasets [4, 5] were evaluated.

In total, 36 wind measurement campaigns located in three different countries and 32 reanalysis datasets were used in this study.

This paper is organized as follows. Data and methodology are described in section 2. The correlations results and its discussion are presented in section 3. Section 4 presents a test case highlighting the impacts of this study achievement. Conclusions are drawn in section 4. The paper finishes with the reference list.

2. Data and methodology

2.1. Local measurements

The masts selected for this investigation are located in Portugal (15), Poland (11) and Romania (10). These meteorological stations were chosen due to the completeness and longevity of the obtained set of records. The wind data are recorded in 10 minutes intervals and the lengths of the datasets vary from 3 to 6 years.

The quality of the selected data is generally very good, with the amount of missing records ranging from 0.4% at PT4 to 16.7% at PT9 (Table 1).

2.2. Reanalysis datasets

Table 2 describes the NCEP/NCAR Reanalysis datasets considered. Datasets were selected with the intent of achieving a adequate geographical coverage of the area were masts are located.

2.3. Methodology

Mainly two aspects of the NCEP/NCAR R2 data sets were tested: i) the correlation with local observations and ii) the inexistence of long-term trends.

The first point intends to address the representativeness of the reanalysis data to describe local wind regimes in the studied areas.

The second intends to rule out any sort of artificial trends in the wind data (mainly wind intensities), that could be present in simulated datasets as this one.

Table 1 – Onsite datasets information

	Met. Station	Datasets extension	Nº of valid records	Wind speed recovery rate (%)	Temporal Resolution
Portugal	PT1	2005 to 2010	203463	96,8	10 min
	PT2	2005 to 2010	135754	86,0	
	PT3	2005 to 2007	151640	96,1	
	PT4	2007 to 2010	209449	99,6	
	PT5	2007 to 2010	197417	93,8	
	PT6	2007 to 2010	201705	95,9	
	PT7	2007 to 2010	197379	93,8	
	PT8	2007 to 2010	196279	93,3	
	PT9	2004 to 2010	306743	83,3	
	PT10	2006 to 2010	242874	92,4	
	PT11	2007 to 2010	198781	94,5	
	PT12	2007 to 2010	191826	91,2	
	PT13	2007 to 2009	147872	93,7	
	PT14	2008 to 2010	157244	99,6	
	PT15	2008 to 2010	148175	93,9	
Poland	PL1	2006 to 2010	246498	93.7	10 min
	PL2	2006 to 2010	232458	88.4	
	PL3	2006 to 2010	240030	91.3	
	PL4	2007 to 2010	186935	88.9	
	PL5	2007 to 2010	195931	93.1	
	PL6	2007 to 2010	189656	90.1	
	PL7	2008 to 2010	152195	96.4	
	PL8	2008 to 2010	142772	90.5	
	PL9	2008 to 2010	148593	94.2	
	PL10	2008 to 2010	149625	94.8	
	PL11	2008 to 2010	146363	92.7	
Romania	RO1	2006 to 2010	239365	91,0	10 min
	RO2	2006 to 2008	154807	98,1	
	RO3	2006 to 2010	221312	84,2	
	RO4	2006 to 2008	133259	84,4	
	RO5	2007 to 2009	137596	87,2	
	RO6	2007 to 2010	199375	94,8	
	RO7	2007 to 2009	152901	96,9	
	RO8	2007 to 2010	184753	87,8	
	RO9	2007 to 2010	190675	90,6	
	RO10	2008 to 2010	143034	90,6	

Table 2 – NCEP/NCAR R2 data considered

Country	NCEP/NCAR grid point		Datasets extension	Vertical level	Temporal Resolution
Portugal	37.5N	-10.0E	1991 to 2010	0.995 Sigma level*	6 hours
		-7.5E			
	40.0N	-10.0E			
		-7.5E			
	42.5N	-10.0E			
		-7.5E			
		-5.0E			
Poland	47.5N	20.0E	1991 to 2010	0.995 Sigma level*	6 hours
		22.5E			
	50.0N	15.0E			
		17.5E			
		20.0E			
		22.5E			
	52.5N	15.0E			
		17.5E			
		20.0E			
		22.5E			
		25.0E			
	55.0N	17.5E			
		20.0E			
		22.5E			
		25.0E			
Romania	42.5N	20.0E	1991 to 2010	0.995 Sigma level*	6 hours
		22.5E			
		25.0E			
		27.5E			
		30.0E			
	42.5N	20.0E			
		22.5E			
		25.0E			
		27.5E			
		30.0E			
	42.5N	20.0E			
		22.5E			
		25.0E			
		27.5E			
		30.0E			

3. Correlation analysis between onsite and reanalysis wind speed data

In order to assess the wind speed variability of a given site, the direct comparison of onsite measured data with reanalysis datasets is widely used in industry.

However there is an initial difficulty when comparing point data (local measurements) with spatial grid data (the reanalysis) which is knowing what is exactly the most representative grid cell from the grid cells surrounding a given mast location.

In Figure 1 each corner of the square represents the centre of a reanalysis grid cell where wind speed over time is known. Each square represents a quarter of the grid cell size and point $P(x_p, y_p)$ marks the location of a given local mast. Should the mast data be correlated with the grid cell with the nearest centre? Should it be correlated with the grid cell with the nearest border?

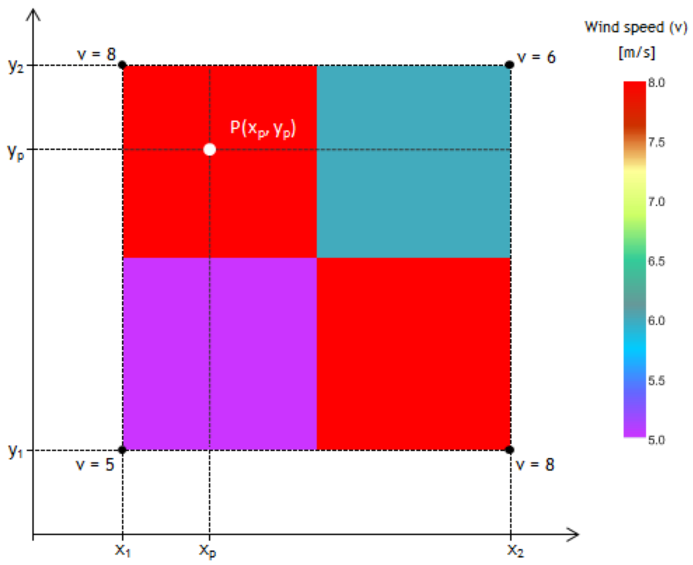

Figure 1 – Schematic illustrating the concept of nearest neighbour

This question can be tested by addressing the correlation for different grid point definitions. By grid point definition one can understand the location, within a reanalysis grid cell, where the grid cells values are considered to be representative. The grid point definitions tested were grid point at the centre of the grid cell and at each corner of the grid cell (bottom left, bottom right, top left, top right).

The differences between the mean coefficients of determination (R), relative to the R for grid point definition, are presented in Table 3.

Table 3 – Differences in R-values when using different grid point definitions relative to grid-cell centre definition

	Central	**Bottom left**	**Bottom right**	**Top left**	**Top right**
Mean	**0,000**	**-0,010**	**-0,038**	**-0,034**	**-0,069**

The results show that all other definitions have slightly lower mean R than grid-cell centre definition. However, differences are quite small, when compared with the variation of the whole 36 point sample itself (see table 4).

Based on this, the grid cell centre was considered to calculate the correlations represented in table 4 under “closest” grid point.

3.1. Interpolation techniques

Two different interpolation schemes, using the four reanalysis grid points closer to each mast, were tested: bilinear and Modified Shepard Method (using Weighted Least Square Interpolation) [6].

Grid points were considered at the grid cell centre, according to the findings from the previous section.

Bilinear interpolation considers the closest 2x2 neighbourhood of known NCEP/NCAR grid values surrounding the unknown point. It then takes a weighted average of these 4 points to get to its final interpolated value. The concept of bilinear interpolation is illustrated in Figure 2.

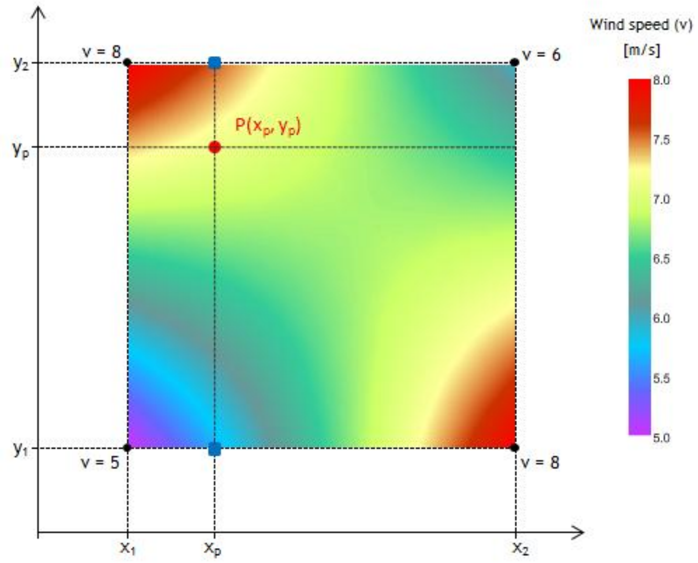

Figure 2 – Schematic illustrating the concept of bilinear interpolation

The desired wind speed quantification is the value v_p at the point (x_p, y_p), which is indicated by the red dot. The approach is to interpolate first in one direction, for example x-direction, to get values at the locations marked by filled blue squares and then interpolate in the other direction to get v_p. Due to the linearity of the method, the order in which interpolations are made in the two directions, x and y, is irrelevant.

In bilinear interpolation scheme, as the interpolating point moves from grid square to grid square, the interpolated function value changes continuously, as can be seen on figure 1. However, the gradient of the interpolated function changes discontinuously at the boundaries of each grid square.

Trying to surpass this limitation, the Modified Shepard Method (using Weighted Least Square Interpolation) was also tested. A brief description of this method is presented below.

Franke and Nielson [7] developed a modification that eliminates the deficiencies of the original Shepard's method [6]. They modified the weight function Wk(x) to have local support and hence to find the overall approximation, and replaced fk with a suitable local approximation Pk(x). This method is called the local modified Shepard method and has the general form,

$$f(x) = \frac{\sum_{k=1}^{n} W_k(x).P_k(x)}{\sum_{k=1}^{n} W_k(x)} \qquad (1)$$

Where, Pk(x) is a local approximant to the function f(x) centred at x(k), The choice for the weight functions Wk(x) used by Renka [8] was suggested by Franke and Nielson [7] and is of the form,

$$W_k(x) = \left[\frac{\left(R_w^{(k)} - d_k(x)\right)}{R_w^{(k)} . d_k(x)}\right]^2 \qquad (2)$$

where dk can be one of two, the Euclidean distance or the Effective distance [9], more advanced formulation, between the points x and x(k), and the constant R(k)w > 0 is a radius of influence about the point x(k) chosen just large enough to include NW points. The data around x(k) only influences f(x) values within this radius.

There are several variations of the original Shepard algorithm based on polynomial (first, second or third degree) or trigonometric functions for Pk. In this work a concept of effective distance was used [8] with a polynomial local approach that use a second degree function, obtained by the weighted least squares method.

In the Figure 3 is presented a wind speed grid based on the same assumptions used for bilinear method. The Modified Shepard Method was tuned to use a concept of effective distance and a second degree polynomial to adjust the points.

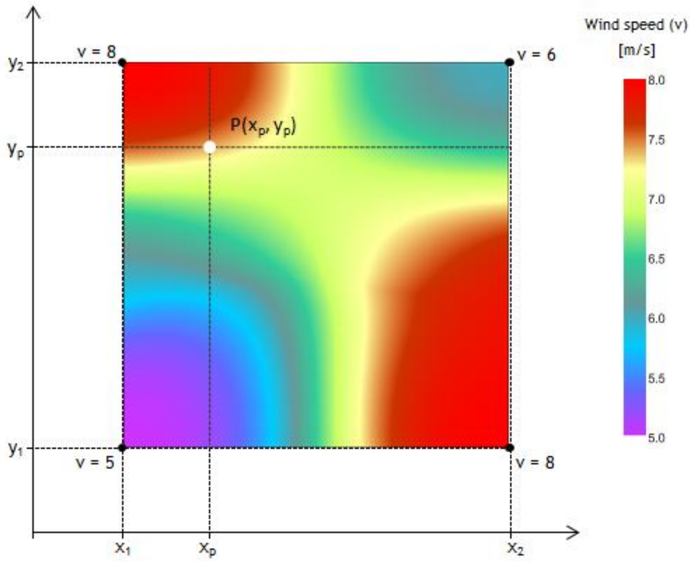

Figure 3 – Schematic illustrating the concept of Modified Shepard Method (with effective distance concept)

In Table 4 the correlation coefficients for the regressions between measurements and both interpolation schemes used are compared with the one obtained considering the “closest” NCEP/NCAR grid point to each meteorological station.

The majority of sites tested show a reasonable correlation between mean monthly wind speeds, regardless of the interpolation scheme used. If we rule out the worst case - PT10 – R value is always above 0,700.

Differences in correlation between interpolation schemes are not representative and well below the variability of the sites’ sample itself. From this, no clear interpolation scheme can be selected, or excluded, yet.

Results are also not very different between regions. Sites in Portugal show slightly worst correlations, but this could easily be explained with the fact that most of those sites’ terrain is more complex than that from the other countries

Table 4 - R-values from the correlation analysis between measurements and different interpolation schemes of the reanalysis datasets to masts location

	R-value using		
	Closest grid point	**Bilinear interpolation**	**Least square interpolation**
PT1	0,732	0,820	0,807
PT2	0,768	0,721	0,754
PT3	0,886	0,894	0,889
PT4	0,845	0,900	0,862
PT5	0,839	0,827	0,837
PT6	0,794	0,788	0,793
PT7	0,827	0,822	0,827
PT8	0,837	0,835	0,838
PT9	0,844	0,819	0,839
PT10	0,598	0,697	0,664
PT11	0,833	0,862	0,846
PT12	0,799	0,826	0,813
PT13	0,884	0,913	0,898
PT14	0,838	0,864	0,844
PT15	0,807	0,833	0,811
PL1	0,820	0,841	0,830
PL2	0,870	0,886	0,876
PL3	0,821	0,837	0,826
PL4	0,910	0,947	0,947
PL5	0,907	0,912	0,907
PL6	0,894	0,897	0,895
PL7	0,925	0,937	0,926
PL8	0,890	0,918	0,908
PL9	0,949	0,949	0,955
PL10	0,939	0,945	0,944
PL11	0,907	0,916	0,912
RO1	0,852	0,855	0,857
RO2	0,852	0,857	0,858
RO3	0,770	0,761	0,768
RO4	0,870	0,872	0,872
RO5	0,874	0,874	0,874
RO6	0,838	0,816	0,830
RO7	0,917	0,894	0,897
RO8	0,861	0,843	0,855
RO9	0,843	0,848	0,848
RO10	0,815	0,810	0,815
Min	**0,598**	**0,697**	**0,664**
Máx	**0,949**	**0,949**	**0,955**
Mean	**0,846**	**0,857**	**0,853**

4. Investigation of the reanalysis wind speed data long-term trends

Besides space continuity, discussed in the previous section, the use of NCEP/NCAR reanalysis data for long term corrections of short term measurements demands a high standard of consistency through time [3, 10, 11].

For all reanalysis datasets used in this work the time consistency was analysed by means of the slope (*a*) of a linear regression (*y*=*ax*+*b*) fitted to all NCEP/NCAR 20 year time series.

Table 5 summarizes what was investigated.

Table 5 – Linear fit slopes of the time evolution of the used reanalysis datasets

	Linear fit slope		
	Portugal	**Poland**	**Romania**
Min	-2.10E-05	-7.00E-05	-3.32E-05
Max	2.88E-05	-8.79E-06	1.10E-05
Mean	6.69E-06	-3.71E-05	-5.58E-06

As it can be seen, no significant strong upward or downward trend was found in any of the data sets used, indicating that the reanalysis data series are consistent through time.

5. Concluding Remarks

Having studied 36 sites in Portugal, Poland and Romania, we can roughly conclude on the benefits of using NCEP/NCAR R2 reanalysis to assess the local long-term wind speed variability.

NCEP/NCAR reanalysis is well correlated with local observation, in terms of monthly mean wind speeds.

Correlations are not very dependent on interpolation schemes and are consistent through the studied sites. If an outlier worst site is excluded, no correlation shows a coefficient of determination (R) below 0,700.

Results are also consistent between the different regions. Sites in Portugal show slightly worst correlations, but this could easily be explained by the fact that most of those sites' terrain is more complex than that from the other countries.

No significant upward or downward trend was found in any of the data sets used, indicating that the reanalysis data series are consistent through time.

With the above, we can assert that NCEP/NCAR R2 reanalysis can be used to assess overall long-term wind speed variability on the studied regions.

Further studies should address the correlation between data with higher time resolution than monthly mean values and also address the representativeness of reanalysis data in regards to specific regions and levels of terrain complexity.

6. References

1 Peter Raftery, Andrew Tindal, Andrew Garrad. *Understanding the risks of financing wind farms*.

2 Paul Gardner et al. *Wind energy - The facts*. A *guide to the technology, economics and future of wind power*. European Wind Energy Association 2009.

3 Sónia Liléo, Olga Petrik. *Investigation on the use of NCEP/NCAR, MERRA and NCEP/CFSR reanalysis data in wind resource analysis*. EWEC 2010 procedings.

4 Kistler R, Kalnay E, et al. *The NCEP/NCAR 50-year reanalysis*. 1999.

5 Kanamitsu M. et al.. NCEP-DEO AMIP-II Reanalysis (R-2). Bulletin American Meteorology Society, 83, 1631-1643, 2002.

6 Shepard, D. (1968), A two dimensional interpolation function for irregularly spaced data, Proc. 23rd Nat. Conf. ACM, p. 517-523.

7 Franke, R., and Nielson, G. (1980), Smooth Interpolation of Large Sets of Scattered Data, International Journal for Numerical Methods in Engineering, v. 15, p. 1691-1704.

8 Renka, R. J. (1988), Multivariate Interpolation of Large Sets of Scattered Data, ACM Transaction on Mathematical Software, v. 14, n. 2, p. 139-148.

9 Lefevre, M., Remund, J., Albuisson, M., and Wald, L. (2002), Study of the effective distances for interpolation schemes in meteorology.

10 Neil Atkinson et al. *Long-term wind speed trends in northwestern europe*.

11 S. C. Pryor, R. J. Barthelmie, J. T. Schoof. *Inter-annual variability indices across europe*. Wiley Interscience, Wind Energy 2006; 9: 27-38.