Renewable and Sustainable Energy Reviews 39 (2014) 509–529

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser

Estimation of wind resources in the coast of Ceará, Brazil, using the linear regression theory

Marcos Antonio Tavares Lira[a,*], Emerson Mariano Da Silva[b], José Maria Brabo Alves[c], Gielson Vitor Oliveira Veras[a]

[a] Department of Electrical Engineering, Federal University of Piauí, Brazil
[b] State University of Ceará, Brazil
[c] Ceará Foundation of Meteorology and Water Resources, Brazil

ARTICLE INFO

Article history:
Received 2 July 2012
Received in revised form
15 May 2014
Accepted 9 July 2014

Keywords:
Wind power
Linear regression
Wind estimation

ABSTRACT

This work is concerned with the estimation of onshore wind resources in the coast of Ceará, Brazil by using the linear regression method. The main focus is to estimate the average wind speed at several altitudes from data collected at the surface. Two specific areas are investigated i.e. Paracuru and Camocim, which are located in the state of Ceará, Brazil. The same methodology is adopted in both cases, where the regions are initially characterized by obtaining the daily and monthly average wind speed profiles from raw data collected by a Platform for Data Collection (PDC) and an Anemometric Tower (AT). Data regarding the prevailing wind direction are also recorded. By using the logarithmic wind profile equation, it is possible to estimate the average wind speeds at altitudes of 20 m, 40 m, and 60 m from data collected at 10 m, as it is possible to determine correlation coefficients between the data and those collected by the AT. Linear regression model is used to estimate the average speed for new altitudes. This procedure is carried out during the calibration and model validation. In both periods, the linear regression model has shown good performance in terms of high level of agreement for the data series, estimated data and related correlation coefficients, and also low error values involving the aforementioned series.

Contents

1. Introduction

The wide access to clean and sustainable energy sources is undoubtedly one of the greatest challenges facing the modern

* Corresponding author. Tel.: +55 86 3237 1565; fax: +55 86 3215 5699.
E-mail address: marcoslira@ufpi.edu.br (M.A.T. Lira).

http://dx.doi.org/10.1016/j.rser.2014.07.097
1364-0321/

world [1]. Nowadays it is possible to notice that several issues regarding energy consumption and generation are widely discussed in media either directly or indirectly.

The encouragement and incentive towards the use of wind energy must be permanently present in public policies related to national energy security. The state of Ceará reached 671 MW of installed wind power generation capacity in 2013 and became the largest wind energy producer in Brazil. Of the total amount, 509.3 MW are produced by 14 wind farms that have been implemented since 2007, while 17.4 MW are from three farms that were installed during the 1990s (Secretary of Infrastructure of Ceará-SEINFRA/CE, 2013).

The energy matrix must be based on the variety of energy sources, which is associated to the complementarity of hydro-environmental systems. This concept depends on the seasonal precipitation cycle, which is directly related to hydropower generation, and also on other atmospheric variables such as the wind speed and solar radiation. For instance, the first semester of the year in northeastern Brazil is known as the rainy period, where higher cloudiness and lower wind speeds are predominant, being the most favorable season for the power generation. When the rainy season ceases or precipitation decreases from June, the use of wind energy is more adequate due to increased wind speeds and solar radiation levels associated to the decreased cloud cover [2,3].

The variability of meteorological parameters is associated to the influence of prevailing weather systems and their performance during the year in a given region depends on local peculiarities of the boundary layer, which cause particular features of wind profiles to change.

In northeastern Brazil, the role of weather systems marks the seasonal wind cycle. On average, from November to January, the frequent incursion of frontal systems moving from sub-tropics of South America to south-central sector of Brazilian northeast region causing rainfall tends to stick with predominant winds blowing from southwest to northeast and in some cases from south to north [4].

In the austral summer and autumn (from January to May), the winds begin to blow climatologically between east and north, with prevailing winds coming from the northeast. During some years, there are different influences on this phenomenon associated with the interannual variability of sea temperature anomalies in the tropical Atlantic Ocean [5], which may modify this regime. During years in which the temperatures of the sea surface (TSS) anomalies in the Tropical North Atlantic are warmer than the TSS anomalies in the Tropical South Atlantic, there is higher frequency of winds blowing from southeast in this part of northeastern Brazil (NEB) [6]. However, within the period from January to May, particularly between March and April in the northern part of northeast Brazil, the Intertropical Convergence Zone (ITCZ) causes more frequent rain events and also weaker winds.

From June to early November, the semi-arid Northeast of Brazil is marked by a period of low rainfall, except for the eastern sector of the region, whose rainy season occurs between April and July. Due to the influence of the so-called wave disturbances from the east, atmospheric disturbances move from the Atlantic Ocean towards the continent, causing rain in the forest zone of northeast Brazil [7].

Dry air masses are predominant between June and November not only in northeast Brazil, but also in most of South America. This phenomenon is associated with the displacement of high pressure system called semi-stationary anticyclone to the east coast and southeastern Brazil, resulting in the predominance of winds blowing from the east-southeast quadrants in northeast Brazil. During the first half of the year, this system lies predominantly in the southeastern sector of the Atlantic Ocean.

In addition to synoptic features that influence the wind regime in northeast Brazil inter annually, local circulations such as land and sea breezes and local effects of planetary boundary layer (which are mainly associated to turbulent processes) are also factors that play an important role in the daily wind direction and speed [8–12]. The land and sea breezes caused by the thermal contrast continent–ocean are responsible for changes in wind direction during the day and night in coastal areas of the northeast. During the day, mainly in the afternoon, when the continent is warmer than the sea, a lower pressure center of the wind is created, which tends to blow into the continent in the form of the sea breeze. Overnight, the characteristics are the opposite, while the sea loses heat more slowly than the continent generating lower pressures and causing winds to blow from the continent towards the ocean in the form of land breeze. Procedures for instability and atmospheric boundary layer characteristics such as roughness and even differences in vegetation and topographical features are also very responsible for diurnal changes in the regime of wind speed and direction [10,11].

In northeast Brazil, the states of Maranhão, Piauí, Ceará, and Rio Grande do Norte are favored by the combination of winds from the east with land and sea breezes, leading to annual average wind speeds within the range from 6 m/s to 9 m/s. The coastline that stretches across the states of Paraíba and Bahia offers speeds from 3.5 m/s to 6 m/s. The areas of hills and plateaus that extend over the Rio Grande do Norte coast to the state of Rio de Janeiro have annual average wind speeds from 6.5 m/s to 8 m/s [20].

Additionally, the production of electricity in 2011 increased by 9.1% and the percentage of renewable energy in the Brazilian energy matrix reached 45.4%. The generation of energy from renewable sources increased by 5%, while wind energy showed the highest growth rate (50.5%) followed by biomass (18.1%). The Brazilian incentive program for alternative energy sources (PROINFA) created in 2002 associated with wind energy auctions promoted by the federal government since 2007 have been responsible for the significant increase of wind energy generation in the energy matrix.

Even in noncoastal areas of the northeast, there are sites whose wind potential must be tapped e.g. Araripe, which is a

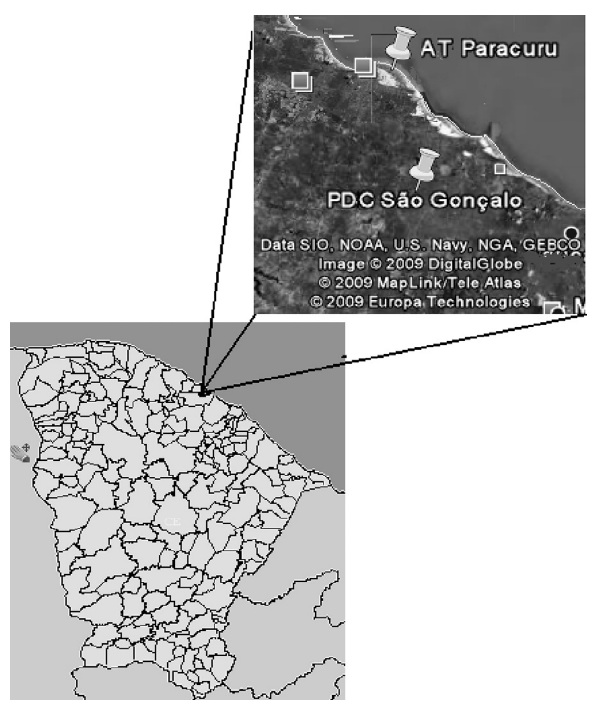

Fig. 1. Map of the state of Ceará highlighting the study area (Paracuru).
Source: Adapted from the Google maps

high altitude region located in the southern state of Piauí supposed to present an installed capacity of 3 MW by 2016.

In order to determine whether a given region represents a potential wind site, it is necessary to acquire reliable data regarding the direction and average wind speed. Several works in literature show that most of wind data are obtained at the surface (10 m), while more reliable data would be obtained precisely at the height where the wind turbine is supposed to be installed [2,3,13]. In order to correct this in practice, it is possible to install ATs responsible for collecting such data during a short interval i.e. two or three years. However, a considerable amount of financial and technical resources is demanded in this case. Based on the aforementioned limitation, this work presents a method for the estimation of wind speed at altitudes where wind generators are installed by using data collected at the surface.

2. Methodology and data

Two distinct geographical regions in the state of Ceará are investigated in this work i.e. Paracuru and Camocim. For geographical location purposes, Figs. 1 and 2 show the state of Ceará map illustrating the study region and its corresponding ATs and PDCs.

The region of Paracuru comprises an AT and a PDC, which are installed at the cities of Paracuru and São Gonçalo do Amarante, respectively. The AT in Paracuru is installed at a point whose geographical coordinates are 03°24′42.4″S and 38°59′02.8″W, while the PDC in São Gonçalo do Amarante is installed at coordinates 03°39′0.8″S and 38°56′32.6″W. The anemograph used in the AT for data collection has a sampling rate of 0.5 Hz (signals every 2 s) and it does record data within the integration interval of 10 min.

The region of Camocim comprises an AT placed at the city of Camocim and a PDC at the city of Barroquinha. The AT is installed at a point whose geographical coordinates are 02°51′56.7″S and 40°53′09.2″W. On the other hand, the PDC in Barroquinha is installed at a point whose coordinates are 2°55′48.2″S and 41°7′7.3″W.

The data regarding the wind speed and direction used in this study are obtained from two sources: PDCs monitored by the Ceará Foundation of Meteorology and Water Resources (CFMWR) and ATs owned by SEINFRA/CE. The ATs are equipped with sensors

Fig. 2. Map of the state of Ceará highlighting the study area (Camocim). Source: Adapted from the Google maps

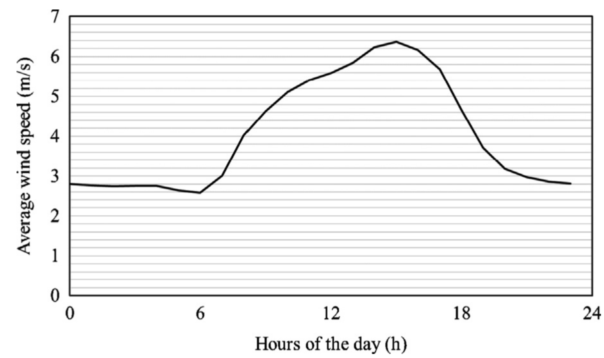

Fig. 3. Profile daily average wind speed in PDC of São Gonçalo do Amarante.

to measure the wind speed at three levels (20 m, 40 m, and 60 m) and also the wind direction at two levels (40 m and 60 m). Besides, the PDCs have sensors for the measurement of the wind speed and direction 10 m above the surface. Data on wind speeds from both the ATs and the PDCs were used to obtain the daily, 5-day, 10-day, and 15-day averages in both study regions.

With the available data regarding wind speed and direction, the first methodology proposed in this paper was carried out. It is worth to mention that some inconsistent values and even non-existent ones were identified in rough data. In order to solve this problem, the gap filling method called cross validation has been adopted [15,19].

The cross validation method is a procedure where each sample is taken from the data set and an interpolation is made to assess its value. The cross validation assumes that a given value is estimated from the data surrounding at a certain point where data have not been collected. After the estimation, the value may be reintroduced in the data series.

Considering the scope of this work, it is necessary to correlate data from PDCs with those from ATs in order to investigate how a given data set explains the variability for another one. In the region of Paracuru, data from the AT in Paracuru was correlated with data from the PDC in São Gonçalo do Amarante. On the other hand, data from the AT in Camocim was correlated with data from the PDC in Barroquinha.

Data regarding the AT in Paracuru and the PDC in São Gonçalo do Amarante were extracted from a series that runs from 08/18/2004 to 09/05/2006, totaling 630 days of observations. This period was divided into two parts and the first 360 days are so-called calibration period, while the remaining 270 days are the model validation period. Analogously, data regarding the AT at Camocim and the PDC at Barroquinha were extracted from a series that runs from 22/8/2004 to 2/12/2006, totaling 540 days of observations. The aforementioned period is also divided into two parts, where the calibration period covers the first 360 days and the validation period comprises the remaining 180 days. The reason we used a shorter time series for this region is based on the fact that the period of observed data is limited.

For both regions, the averages were calculated from data sets of wind speed regarding PDCs and ATs for four different intervals (daily, 5-day, 10-day, and 15-day periods). Then, by adopting PDC data as reference, the expression of the logarithmic wind profile can be used as

$$v(z) = v(z_R)\frac{\ln(z/z_0)}{\ln(z_R/z_0)} \tag{1}$$

where $v(Z_R)$ is the average reference speed at height Z_R, and Z_0 is the surface roughness length

Eq. (1) is used to estimate values for the wind speeds at 20 m, 40 m, and 60 m high and to establish a proper comparison with data collected from the respective AT. It is noteworthy that the aforementioned expression uses $Z_0 = 0.0002$ m, which is a value typically adopted in literature for coastal regions [13,14].

After the calibration period, the correlation for three levels in height was investigated, which involves parameters estimated from the data surface and data collected by the PDCs and ATs. Each scatter plot constructed during the calibration period leads to the equation of a line, as well as their respective determination coefficients. Then, the linear regression model is applied by adjusting the data so that each ordered pair (AT observed value, estimated PDC value) matches its respective line. For each equation in the form $y = \beta_0 + \beta_1 x$, and by applying the values of y (obtained from the use of PDC data in Eq. (1)), new values for x (wind speeds observed in ATs) or x' (wind speeds estimated by ATs) are obtained.

It is then necessary to determine the relationship between x and x'. The scan comprises the measurement of correlation and determination coefficients, and also the errors expressed in terms of the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and percent error (ε) between the estimated and observed values. Parameters RMSE, MAE, and ε are defined as

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (p_i - o_i)^2}{n}} \tag{2}$$

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} p_i - o_i}{n} \tag{3}$$

$$\varepsilon = 100\% \times \frac{(p_i - o_i)}{o_i} \tag{4}$$

where p_i represents the estimated value, o_i is the observed value, and n indicates the number of observations.

According to Zacharias, Heatwole and Coakley (1996) *apud* [15], the RMSE measures the variation of the estimated values around the measured values. On the other hand, MAE indicates the deviation of the mean absolute predicted values in relation to the observed ones. Finally, ε represents the percent deviation between the actual measured value and the inferred one. Ideally, the values for MAE, RMSE, and ε must be close to zero.

For the validation period, linear regression was applied by adopting the same procedure used for the calibration period, i.e. applying the values of y (obtained from the application of PDC data in Eq. (1)) and obtaining x'. Once again, it is possible to state a relationship between x and x'. Besides, new values for the errors and the correlation coefficients are determined during the scan procedure.

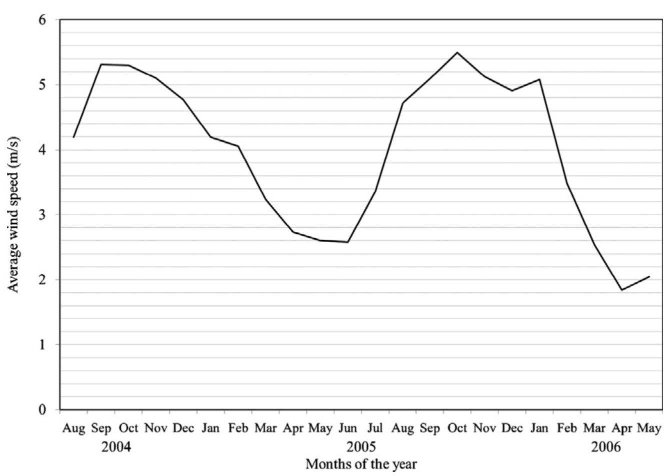

Fig. 4. Profile monthly average wind speed in PDC of São Gonçalo do Amarante.

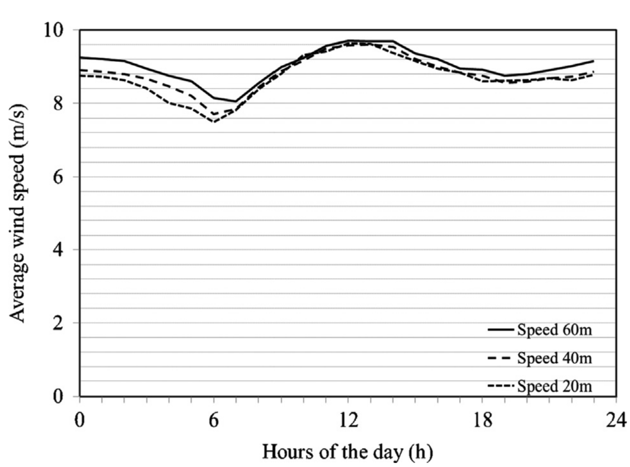

Fig. 5. Profile daily average wind speed in AT of Paracuru.

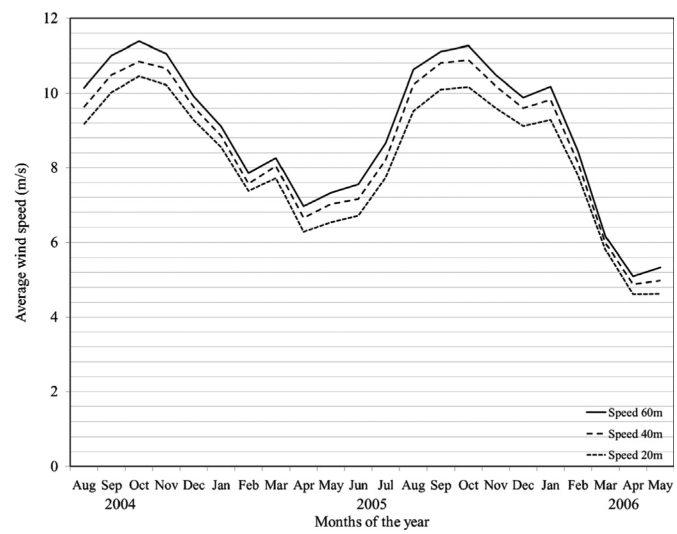

Fig. 6. Profile monthly average wind speed in AT of Paracuru.

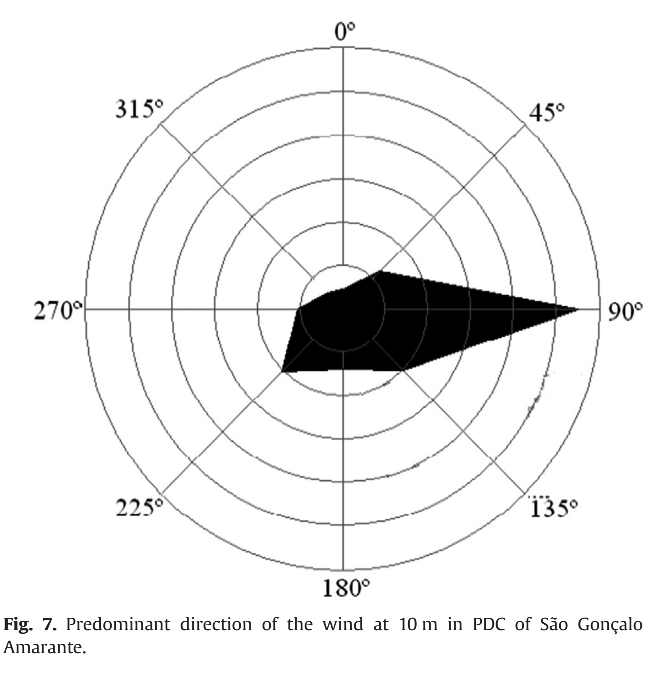

Fig. 7. Predominant direction of the wind at 10 m in PDC of São Gonçalo do Amarante.

3. Results and discussions

3.1. Results for the region of Paracuru

3.1.1. Characterization of the wind profile in the region of Paracuru

Figs. 3 and 4 show the daily and monthly profiles of the average wind speed, respectively, which are obtained from data collected by the PDC in São Gonçalo do Amarante.

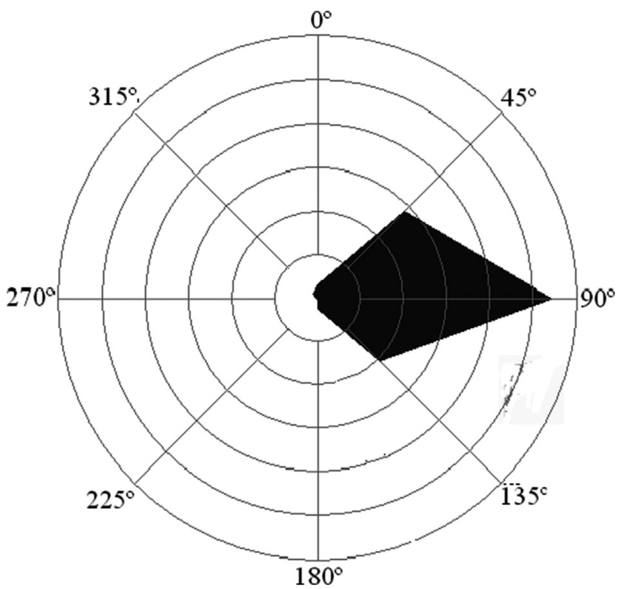

Fig. 8. Predominant direction of the wind at 60 m in AT of Paracuru.

Regarding the daily profile in the region of São Gonçalo do Amarante, it is possible to observe that the highest values for the average wind speed occur during the daytime between 06:00 and 18:00 local time, while the maximum value is 6.3 m/s at about 15:00 local time. The higher intensity of the average wind speed during the daytime can be explained by the presence of sea breeze [16,17].

For the monthly profile, the highest values for the average speed occur in the period from July to December, which match the dry season of the region. Additionally, the lowest average wind speed occurs in the period from January to May during the rainy season, thus validating the results presented by [18].

Considering that the latest wind turbines installed in the state of Ceará are model E70 1.8 manufactured by Wobben (source: Windpower Wobben), whose entry speed is 2.5 m/s (i.e. value from which it is possible to generate electricity), it becomes evident that the average wind speeds shown in the daily and monthly profiles are significant for wind energy generation.

Figs. 5 and 6 show the daily and monthly profiles of the average wind speed obtained from data collected by the AT in Paracuru, respectively. The results show that higher average wind speeds occur as the altitude increases when compared with data from the PDC in São Gonçalo do Amarante.

For the daily profile of the average wind speed, a well-established diurnal cycle exists as in the case of the PDC in São Gonçalo do Amarante, while the maximum average wind speed occurs at about 15:00 local time.

For the monthly profile, the periods where the highest and the lowest average wind speeds occur are similar to the aforementioned ones for the PDC in São Gonçalo do Amarante, thus characterizing the region of Paracuru as a potential site for wind energy generation in the state of Ceará.

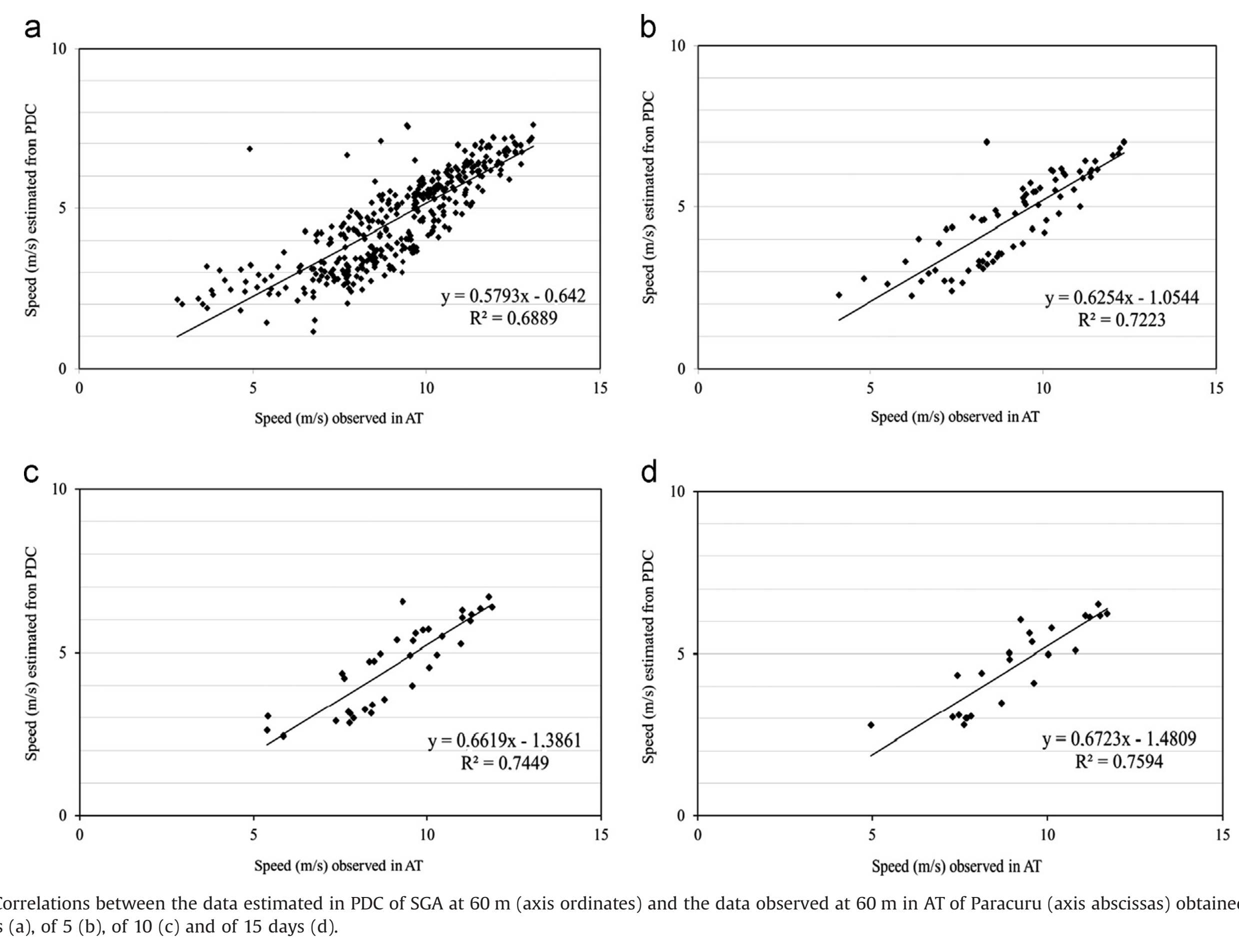

Fig. 9. Correlations between the data estimated in PDC of SGA at 60 m (axis ordinates) and the data observed at 60 m in AT of Paracuru (axis abscissas) obtained for daily averages (a), of 5 (b), of 10 (c) and of 15 days (d).

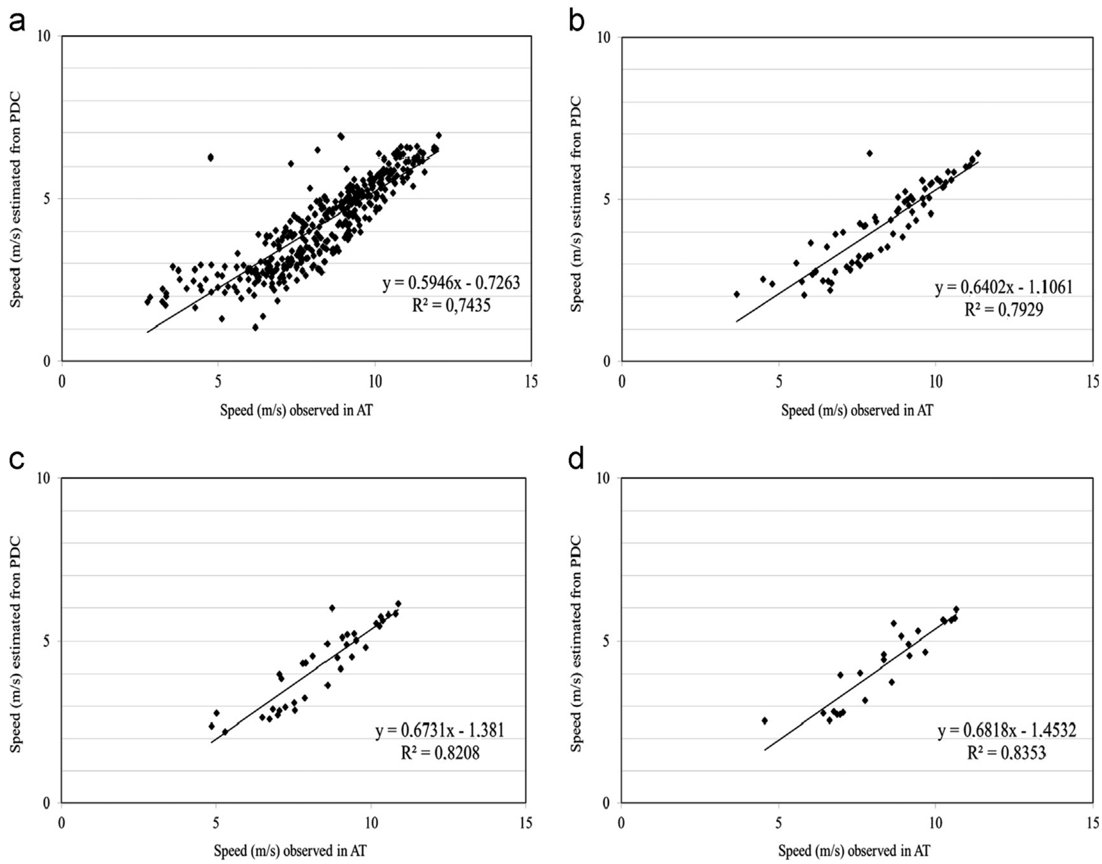

Fig. 10. Correlations between the data estimated in PDC of SGA at 40 m (axis ordinates) and the data observed at 40 m in AT of Paracuru (axis abscissas) obtained for daily averages (a), of 5 (b), of 10 (c) and of 15 days (d).

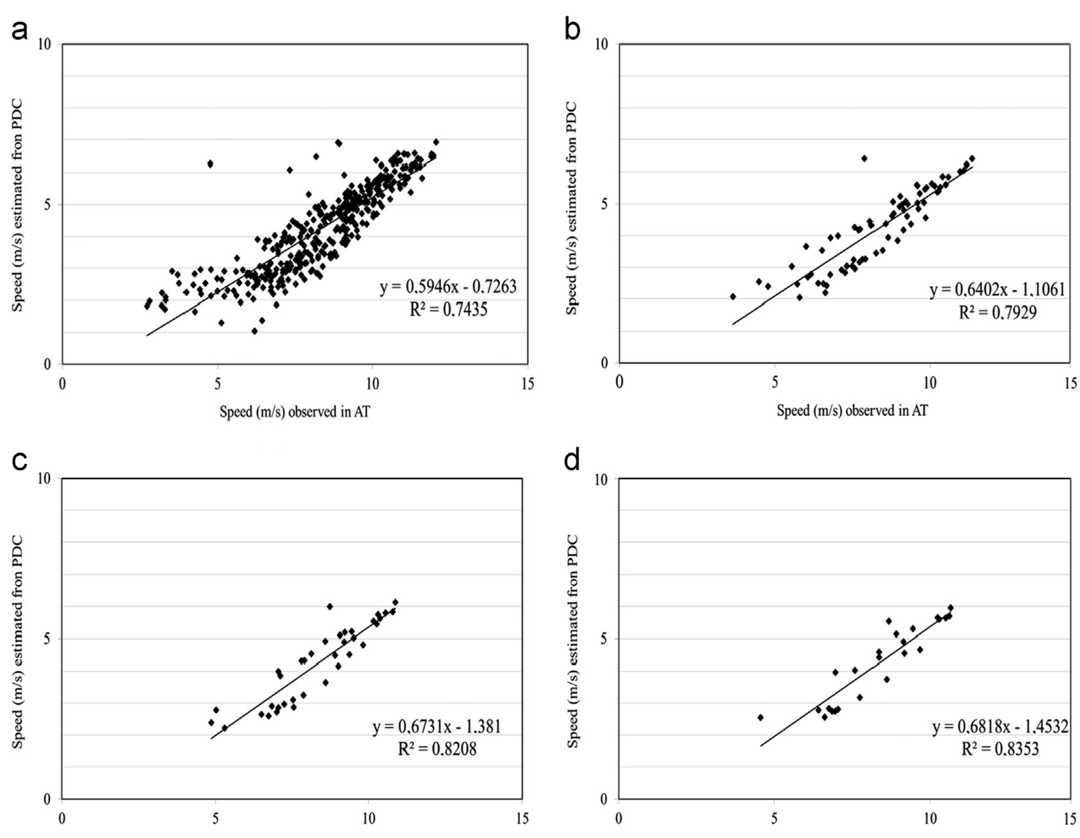

Fig. 11. Correlations between the data estimated in PDC of SGA at 20 m (axis ordinates) and the data observed at 20 m in AT of Paracuru (axis abscissas) obtained for daily averages (a), of 5 (b), of 10 (c) and of 15 days (d).

The profiles also show that at a given altitude there are daily average wind speeds that reach about 10.0 m/s, while the monthly average values exceed 4.0 m/s even in the rainy season, which is higher than the entry speed of the wind turbine.

Figs. 7 and 8 show the predominant wind direction in the regions that correspond to the PDC in São Gonçalo do Amarante and the AT in Paracuru, respectively.

It is observed in Fig. 7 that the predominant wind direction in São Gonçalo do Amarante is from east, with variations over the year between the southwest and northeast directions. Fig. 8 shows that the prevailing wind direction in Paracuru is from east, varying between southeast and northeast directions over the year. Thus, the influence of the trade winds in the region is evident, as mentioned in [16].

3.1.2. Calibration period – region of Paracuru

Figs. 9, 10, and 11 show the correlations between the average wind speed and the estimated values for 60 m, 40 m, and 20 m, respectively, where data collected in the surface by the PDC in São

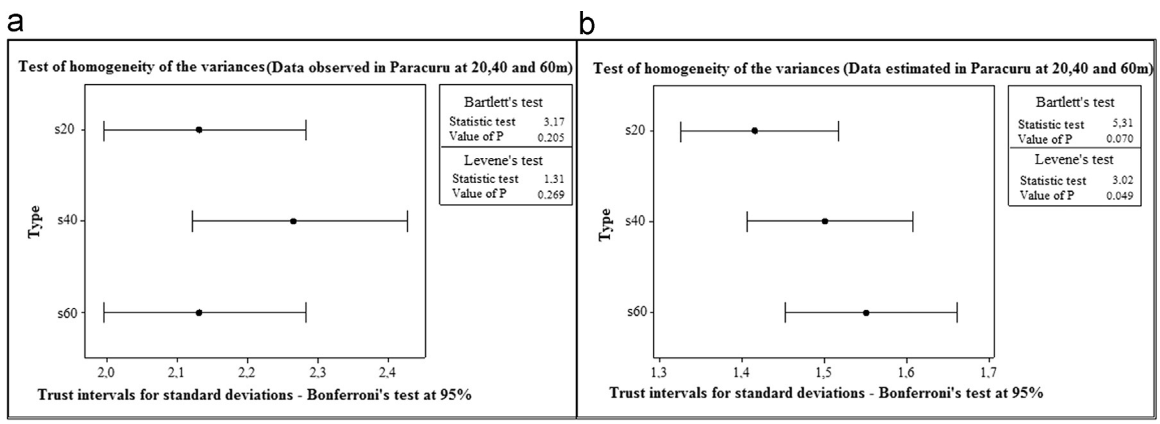

Fig. 12. Test of homogeneity of the variances in region of Paracuru: data observed (a) and data estimated (b).

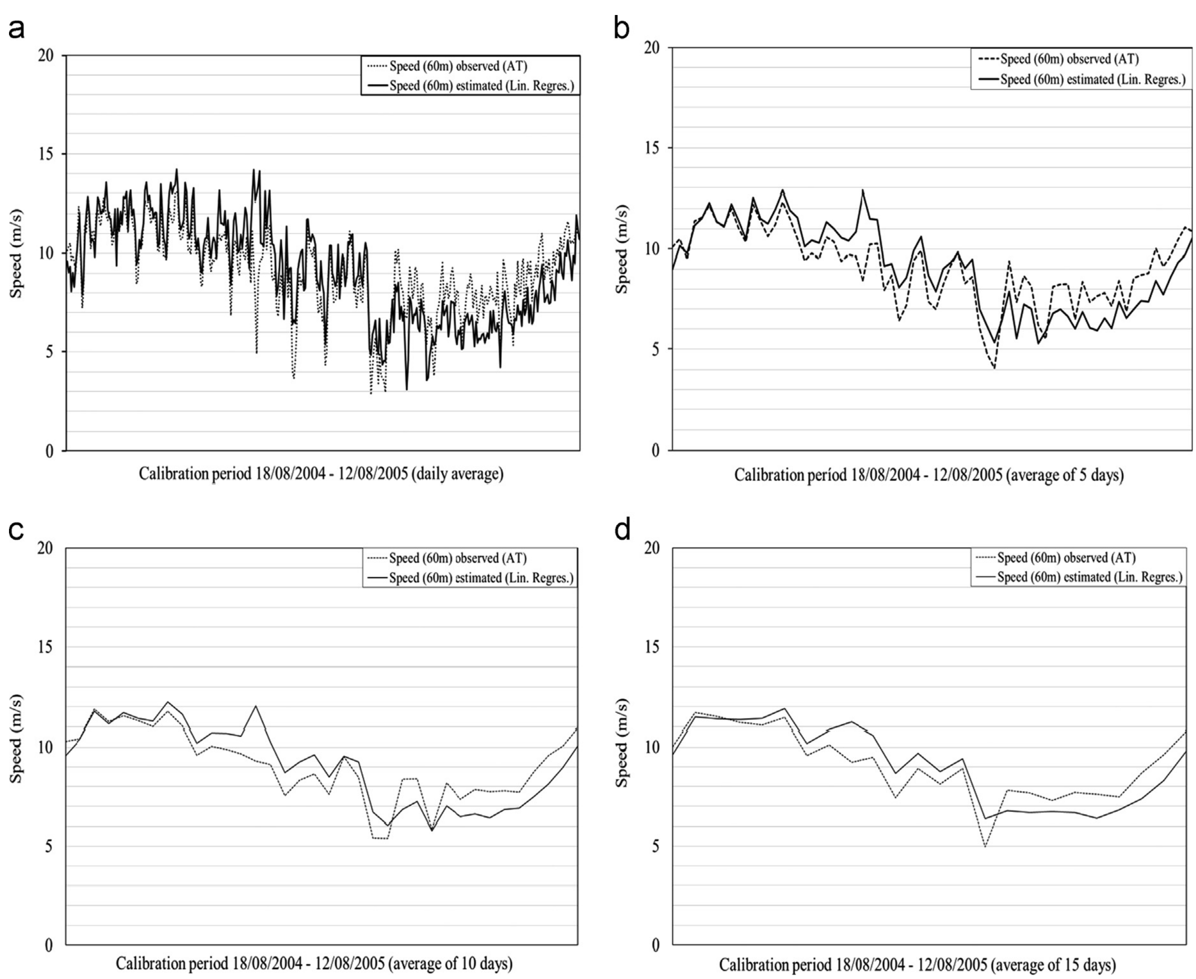

Fig. 13. Comparison between series of data estimated for the lineal regression and data observed in AT of Paracuru for the height of 60 m in the calibration period, for daily averages (a), of 5 (b), of 10 (c) and of 15 days (d).

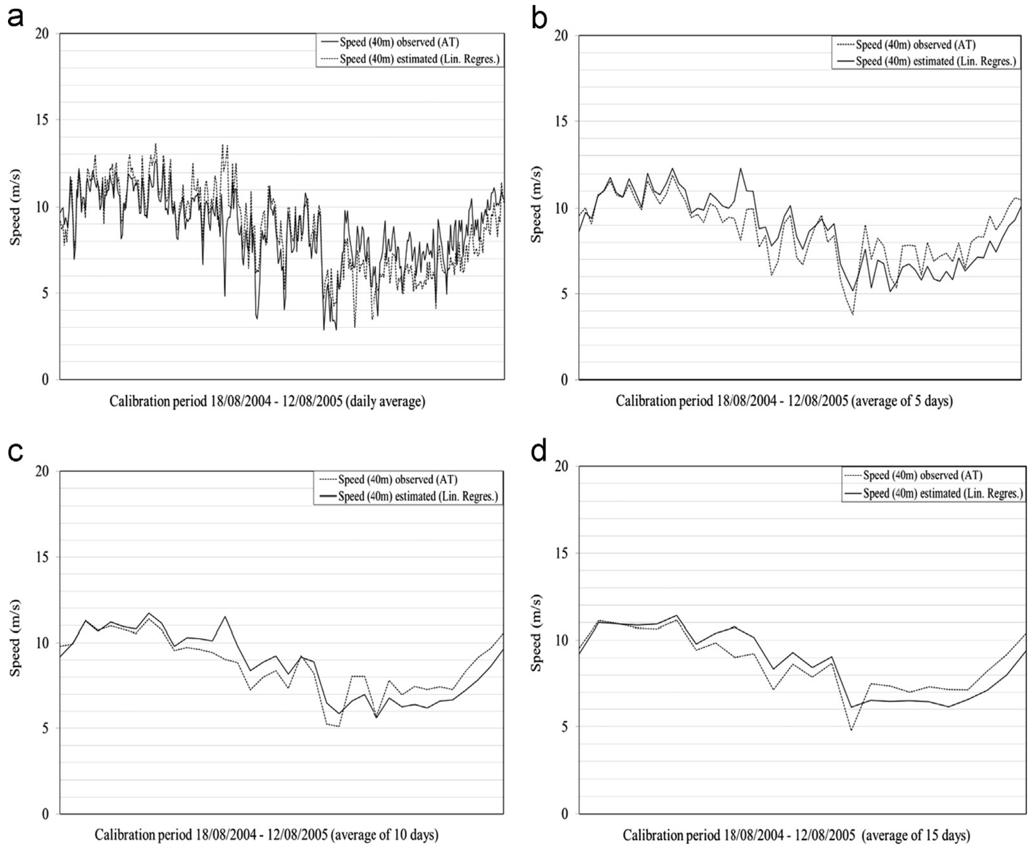

Fig. 14. Comparison between series of data estimated for the lineal regression and data observed in AT of Paracuru for the height of 40 m in the calibration period, for daily averages (a), of 5 (b), of 10 (c) and of 15 days (d).

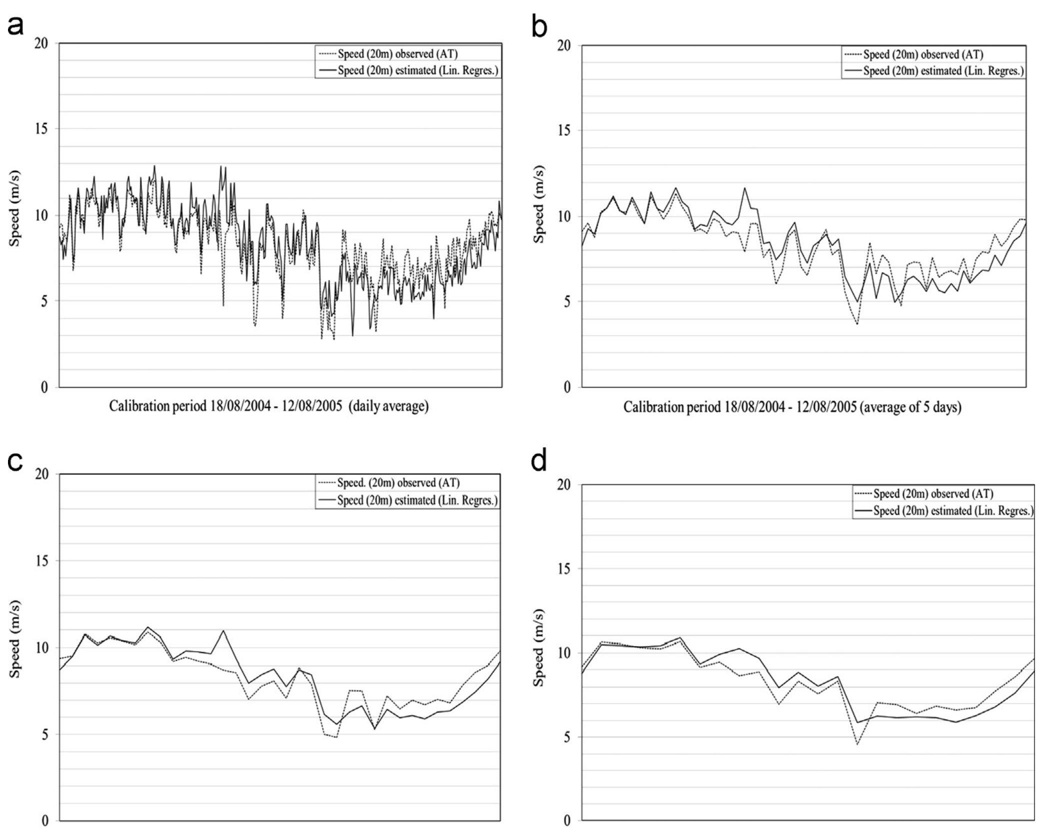

Fig. 15. Comparison between series of data estimated for the lineal regression and data observed in AT of Paracuru for the height of 20 m in the calibration period, for daily averages (a), of 5 (b), of 10 (c) and of 15 days (d).

Gonçalo do Amarante and the AT in Paracuru are represented. In addition, the lines obtained with the linear regression method and also the determination coefficients (R^2) corresponding to the daily, 5-day, 10-day, and 15-day averages are also presented.

The figures also show that the value of R^2 increases when the calculation period for the average wind speed is as extended. For the daily wind speed in Fig. 9a, the linear regression model explains about 69% of the variability involving the observed data,

Table 1
Values of the errors and correlation and determination coefficients between the data estimated in AT for the equations of the regression line and the data observed in AT in the calibration period for the region of Paracuru.

Height (m)	Periods which it was obtained the averages of the speeds of the wind	RMSE (m/s)	MAE (m/s)	ε (%)	r	R^2
20 m	Daily	1.12	0.86	0.10	0.86	0.74
	5 Days	0.87	0.68	0.02	0.89	0.79
	10 Days	0.75	0.60	0.02	0.91	0.83
	15 Days	0.70	0.58	0.11	0.91	0.83
40 m	Daily	1.29	1.01	0.04	0.84	0.71
	5 Days	1.02	0.82	0.04	0.87	0.76
	10 Days	0.88	0.73	0.05	0.88	0.77
	15 Days	0.83	0.71	0.01	0.89	0.79
60 m	Daily	1.39	1.10	0.06	0.83	0.69
	5 Days	1.13	0.92	0.06	0.85	0.72
	10 Days	0.99	0.84	0.13	0.86	0.74
	15 Days	0.94	0.82	0.02	0.87	0.76

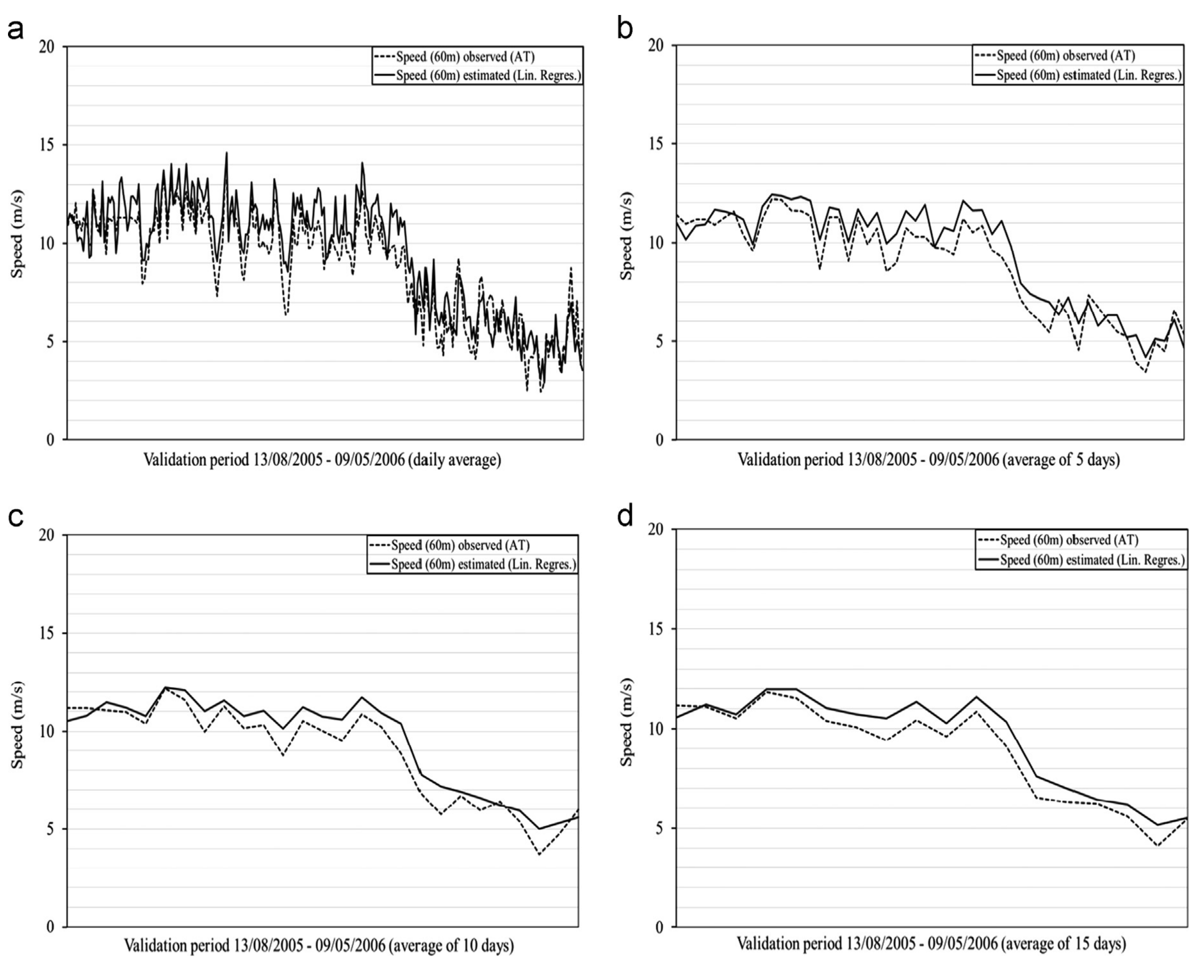

Fig. 16. Comparison between series of data estimated for the lineal regression and data observed in AT of Paracuru for the height of 60 m in the validation period, for daily averages (a), of 5 (b), of 10 (c) and of 15 days (d).

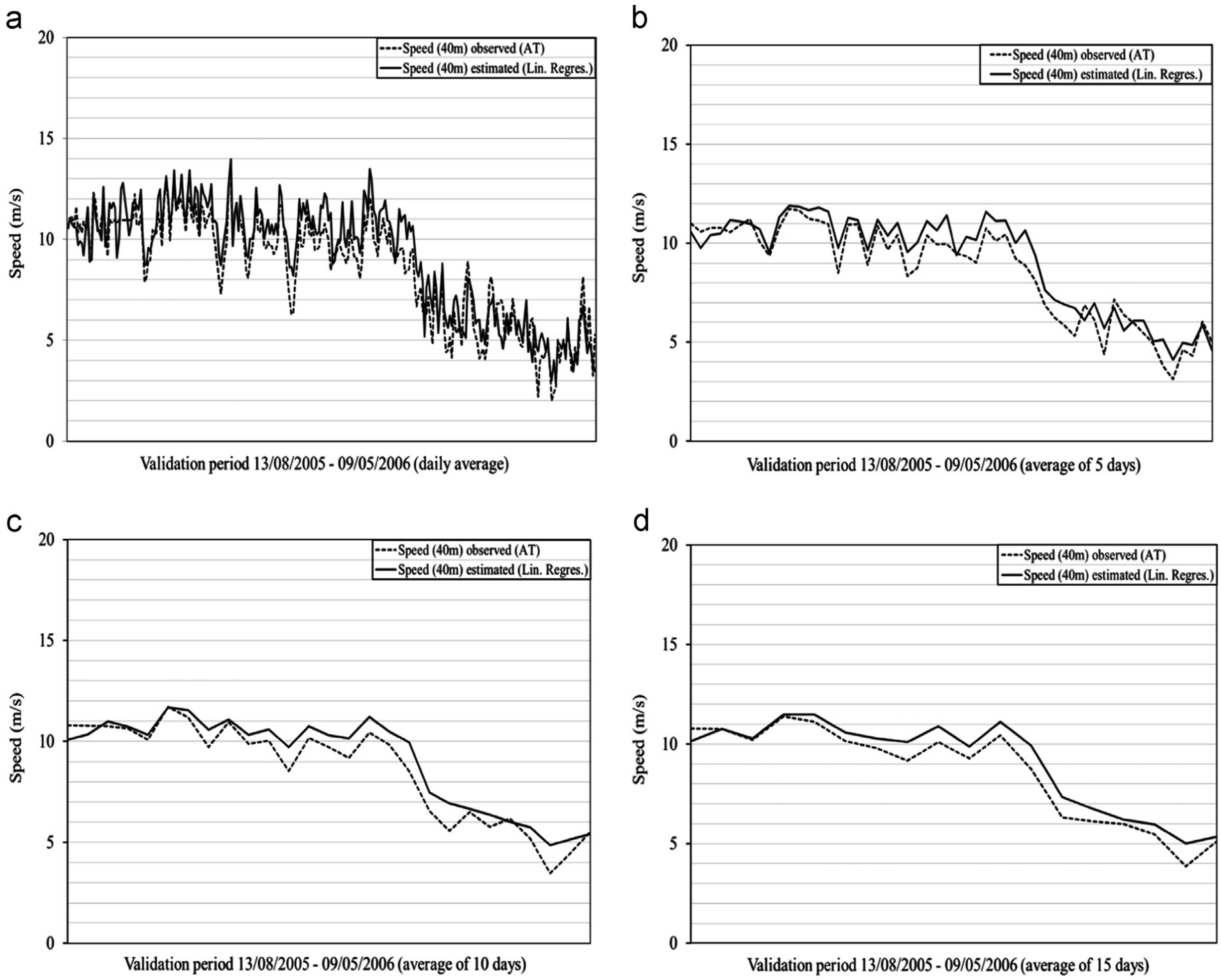

Fig. 17. Comparison between series of data estimated for the lineal regression and data observed in AT of Paracuru for the height of 40 m in the validation period, for daily averages (a), of 5 (b), of 10 (c) and of 15 days (d).

while in Fig. 9d the model explains approximately 76% of such variability for the 15-day average.

Fig. 10 shows that the model explains about 70% of the variability for the observed data when using daily averages, and this percentage reaches values about 79% for the 15 day-average (Fig. 10d).

On the other hand, Fig. 11 shows that the model explains 74% (a) and 84% (d) of the variability for the observed data when the periods for the calculation of averages are one and 15 days.

In order to verify if the variances between AT samples and PDC estimations are homogeneous, a test of homogeneity involving the variance was carried out using software Minitab® version 15 (trial version). Fig. 12 shows the obtained results by using two methods: Bartlett and Levene [19]. Moreover, this figure shows a confidence interval for standard deviations according to Bonferroni's test [19]. Even though $P>0.05$ is recommended, Levene's test has led to $P=0.049$ for the estimated sample. In this case, $P\cong 0.05$ can be fairly adopted, so that hypothesis H_0 is not rejected i.e. the variance of samples are homogeneous.

After obtaining the lines representing the fit of linear regression between the datasets of wind speed, as this procedure is called calibration. Estimated values from such lines were obtained for the three heights measured by the AT in Paracuru, which are then compared with data sets observed in the AT. It is then possible to assess the performance of the linear regression model when reproducing the variability of the average wind speed observed in the region of Paracuru from average speeds estimated by logarithmic wind profile, which are obtained from the wind speed data observed by the PDC in São Gonçalo do Amarante.

Fig. 13 shows the comparison between the data series estimated by the linear regression method and those observed by the AT in Paracuru for daily, 5-day, 10-day, and 15-day averages at 60 m. It is also possible to state that there is an agreement between the values for observed and predicted average speeds during all periods. In general, the data series obtained by linear regression model overestimates the observed data in the dry season and underestimates them during the rainy season. This observation is more evident when increasing the period of calculation for the average wind speeds.

Fig. 14 presents a comparison between the data series estimated by linear regression method and those observed by the AT in Paracuru for daily, 5-day, 10-day, and 15-day averages at the height of 40 m. Once again, the results demonstrate the underestimation and overestimation of periods analogously to those existing in the case of 60 m.

Fig. 15 shows the comparison between the data series estimated by linear regression method and those observed by the AT in Paracuru for average daily, 5-day, 10-day, and 15-day averages at the height of 20 m. In this case, the observed values are also overestimated and underestimated in the dry and rainy periods, respectively, as it occurs for 60 and 40 m.

In all cases, this is because the data sets of observed wind speeds shown in Figs. 5 and 6 have similar temporal variability,

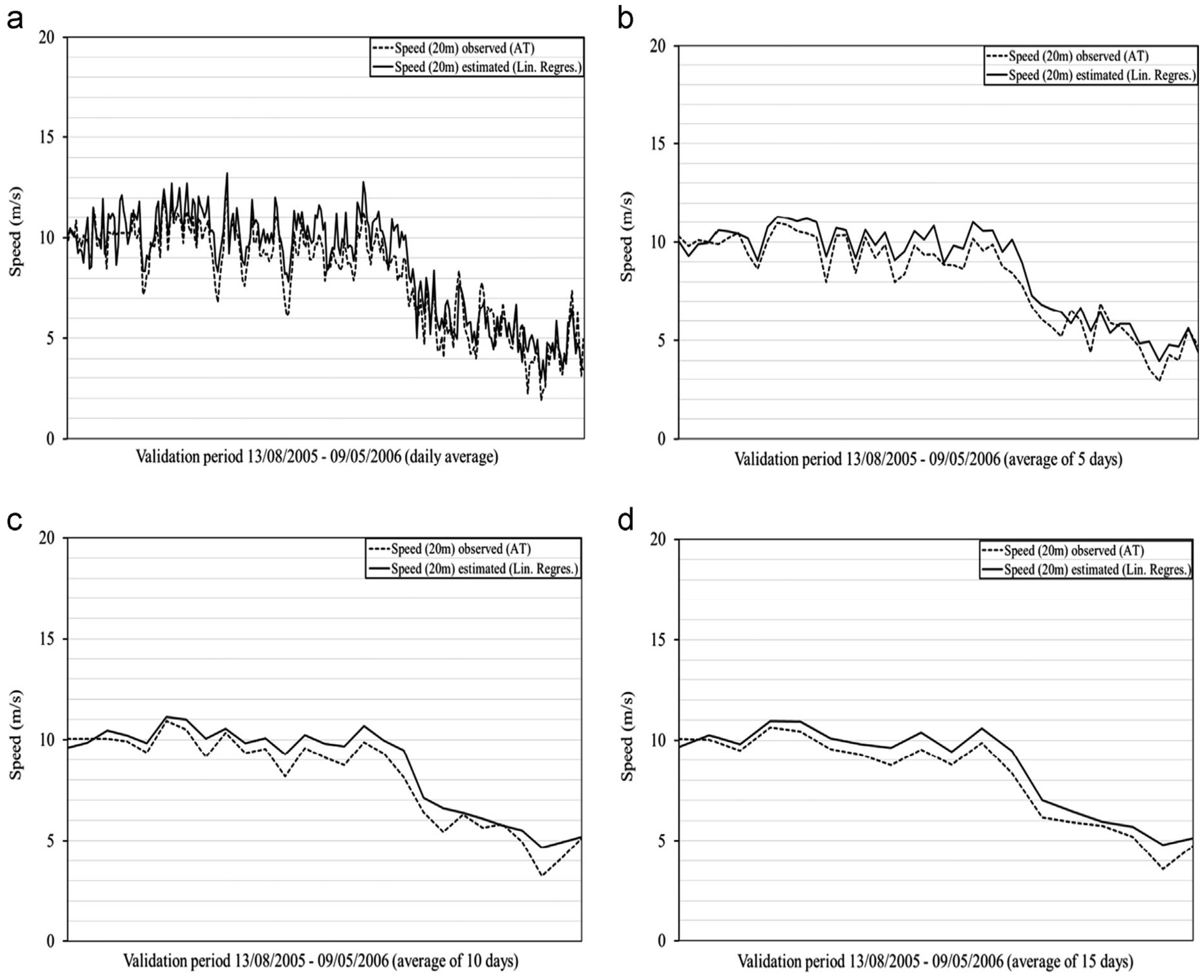

Fig. 18. Comparison between series of data estimated for the lineal regression and data observed in AT of Paracuru for the height of 20 m in the validation period, for daily averages (a), of 5 (b), of 10 (c) and of 15 days (d).

Table 2
Values of the errors and correlation and determination coefficients between the data estimated in AT for the equations of the regression line and the data observed in AT in the validation period for the region of Paracuru.

Height (m)	Periods which it was obtained the averages of the speeds of the wind	RMSE (m/s)	MAE (m/s)	ε (%)	R	R^2
20 m	Daily	1.02	0.85	6.8	0.95	0.90
	5 Days	0.77	0.67	6.4	0.97	0.94
	10 Days	0.70	0.60	6.3	0.98	0.96
	15 Days	0.66	0.59	6.3	0.99	0.98
40 m	Daily	1.07	0.89	5.9	0.94	0.88
	5 Days	0.80	0.69	5.6	0.97	0.94
	10 Days	0.72	0.60	5.4	0.98	0.96
	15 Days	0.66	0.56	5.3	0.98	0.96
60 m	Daily	1.17	0.99	6.6	0.94	0.88
	5 Days	0.88	0.77	6.1	0.97	0.94
	10 Days	0.79	0.68	6.0	0.97	0.94
	15 Days	0.72	0.62	5.8	0.98	0.96

differing only in values i.e. 60 m represents a higher altitude than 40 m, and so on.

Table 1 shows the quantification of errors and the correlation and determination coefficients for the calibration period. In general, for three heights, the lowest value for RMSE, MAE, and ε is verified in the comparison between sets of observed and estimated data during the 15-day average, which also presents the highest correlation and determination coefficients.

In order to verify the fit between the two series, Kolmogorov–Smirnov's test was performed [21], which has shown significance of 95%.

3.1.3. Validation period – region of Paracuru

In this section, the results for the validation period are presented in order to investigate whether the linear regression model has satisfactory performance in the estimation of the average wind speed in the region of Paracuru.

Fig. 16 represents the comparison between the data series estimated by linear regression method and those observed by the AT in Paracuru for daily, 5-day, 10-day, and 15-day averages for the validation period. It seems that there is an agreement among the estimated and observed data series in AT of Paracuru during the validation period, where the overestimation of data obtained by linear regression competed with the observed data in AT is predominant. This is similar to the results found during the calibration period when considering the same time period for

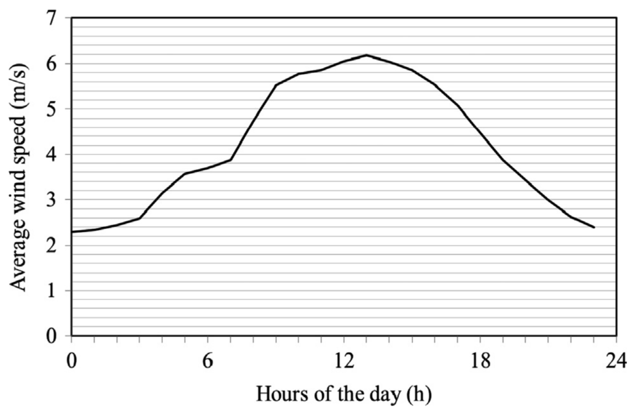

Fig. 19. Profile daily average wind speed in PDC of Barroquinha.

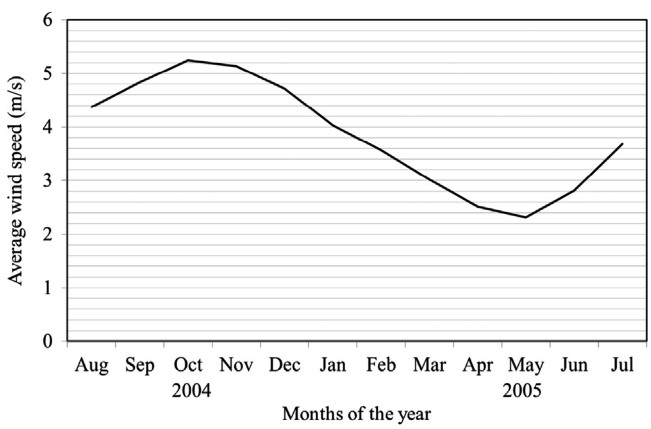

Fig. 20. Profile monthly average wind speed in PDC of Barroquinha.

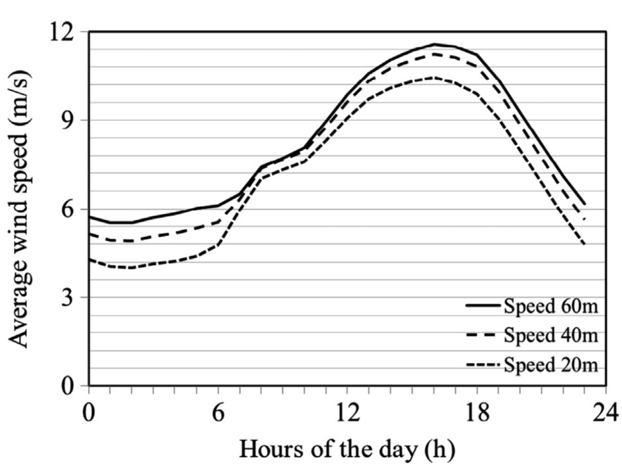

Fig. 21. Profile daily average wind speed in AT of Camocim.

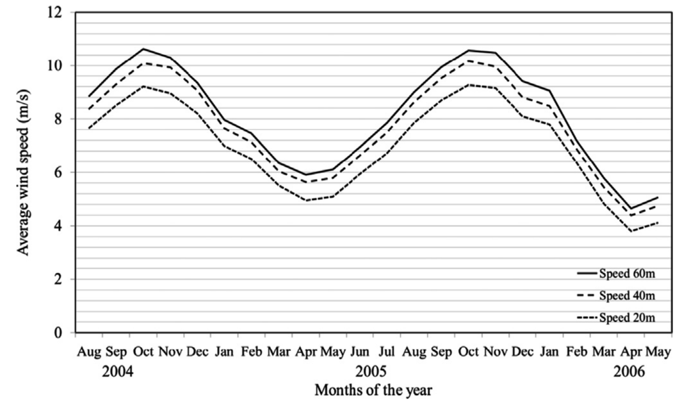

Fig. 22. Profile monthly average wind speed in AT of Camocim.

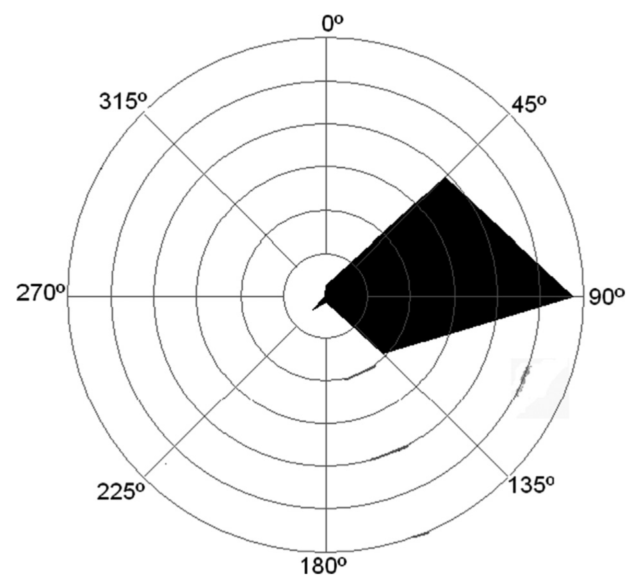

Fig. 23. Predominant direction of the wind at 10 m in PDC of Barroquinha.

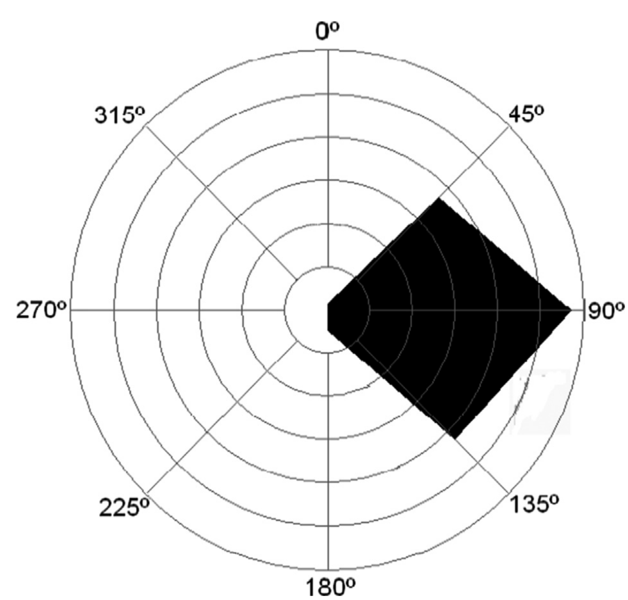

Fig. 24. Predominant direction of the wind at 60 m in AT of Camocim.

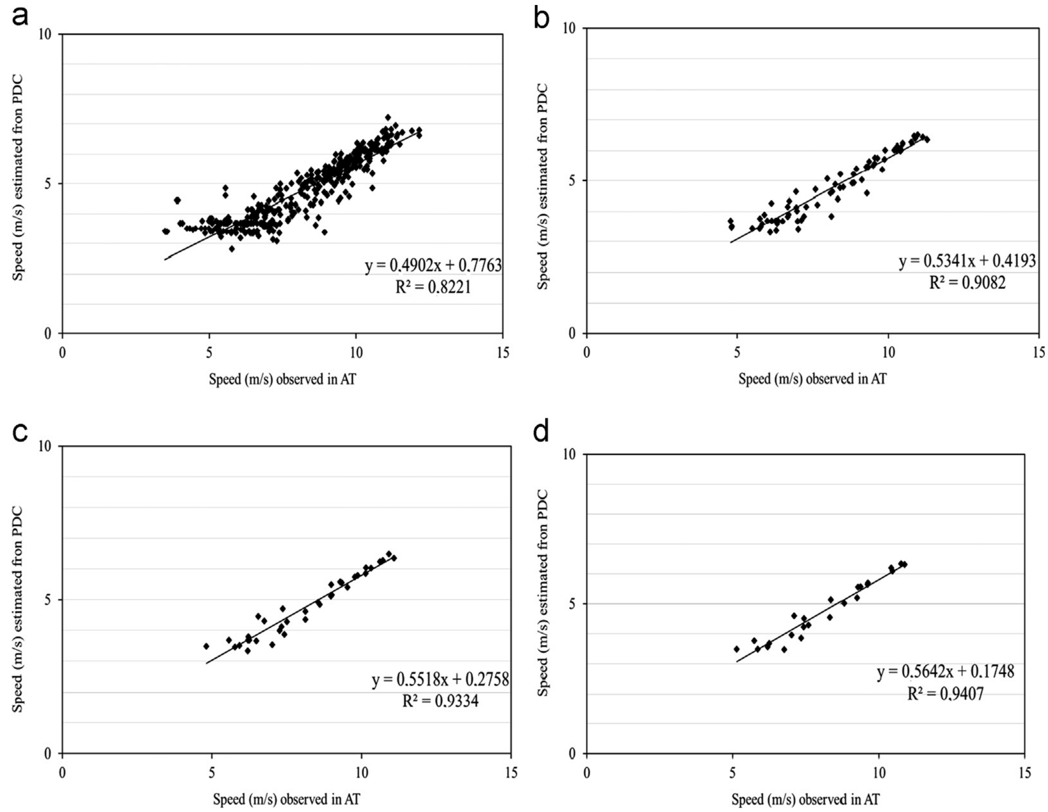

Fig. 25. Correlations between the data estimated in PDC of BAR at 60 m (axis ordinates) and the data observed at 60 m in AT of Camocim (axis abscissas) obtained for daily averages (a), of 5 (b), of 10 (c) and of 15 days (d).

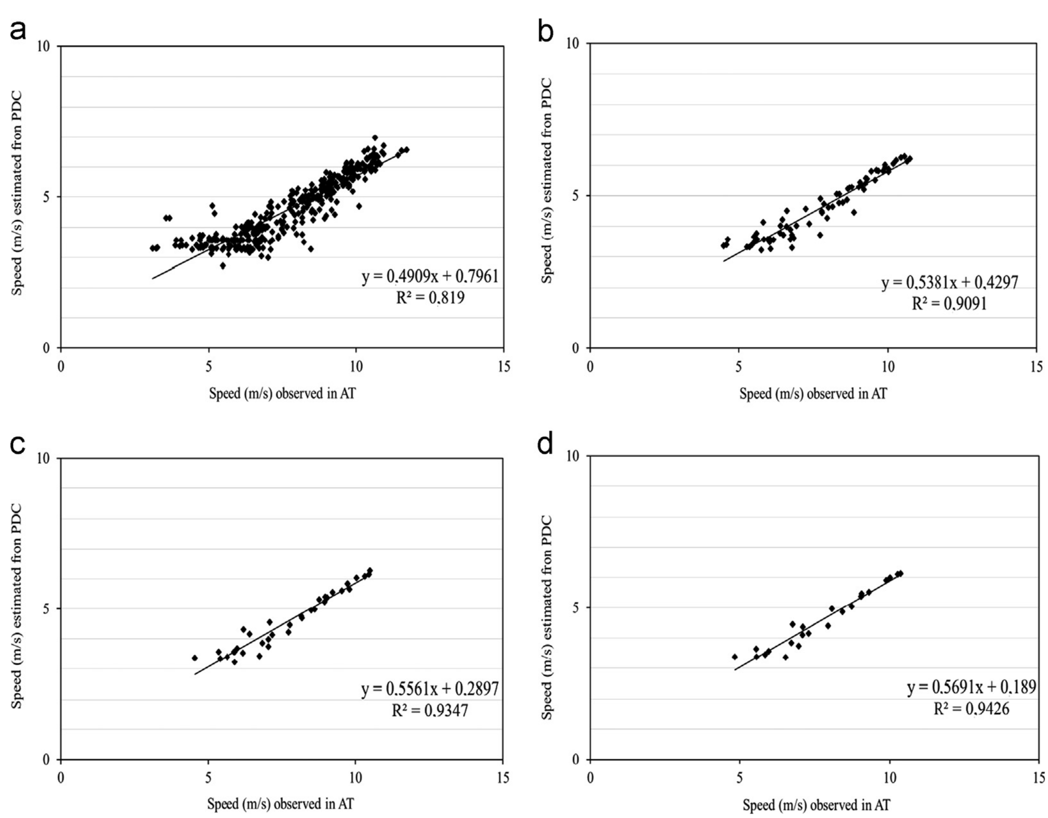

Fig. 26. Correlations between the data estimated in PDC of BAR at 40 m (axis ordinates) and the data observed at 40 m in AT of Camocim (axis abscissas) obtained for daily averages (a), of 5 (b), of 10 (c) and of 15 days (d).

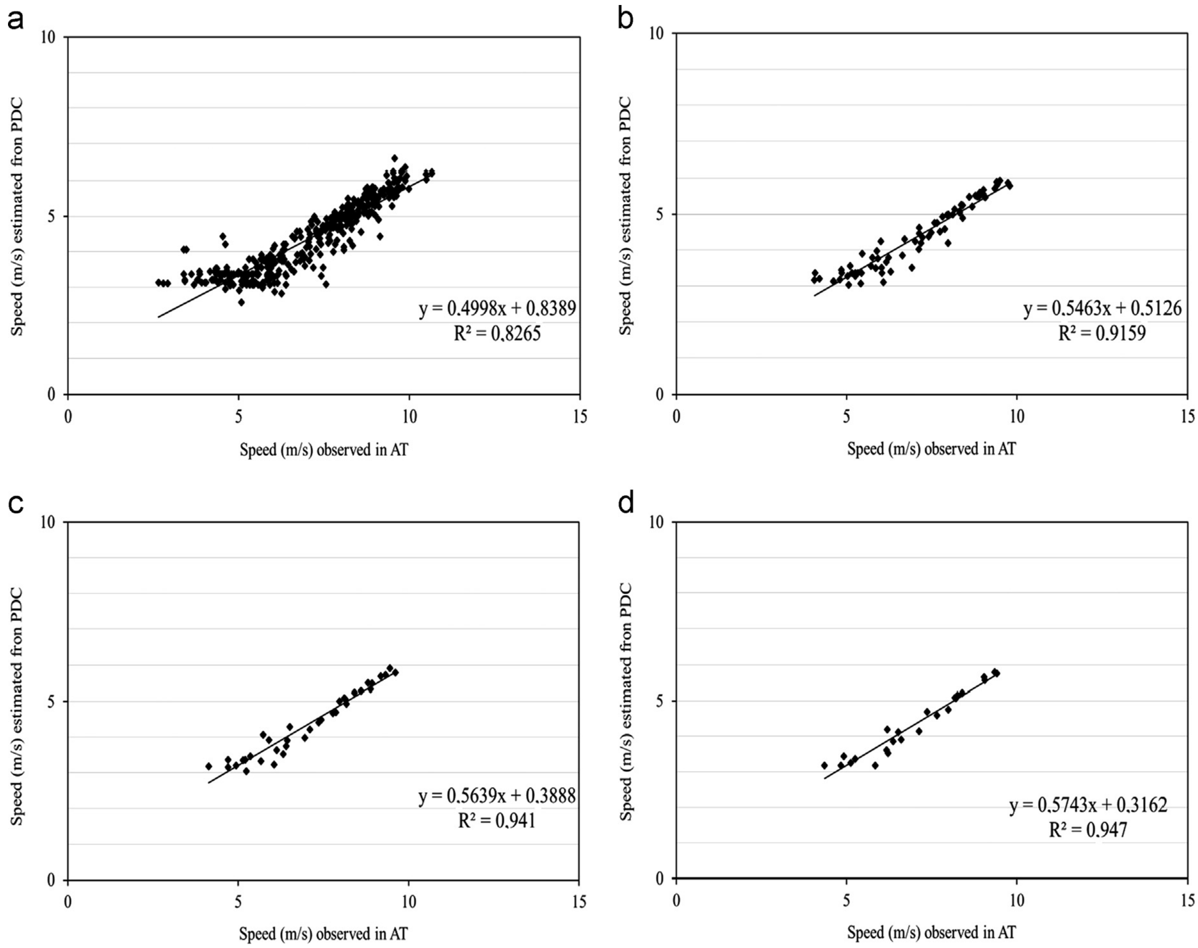

Fig. 27. Correlations between the data estimated in PDC of BAR at 20 m (axis ordinates) and the data observed at 20 m in AT of Camocim (axis abscissas) obtained for daily averages (a), of 5 (b), of 10 (c) and of 15 days (d).

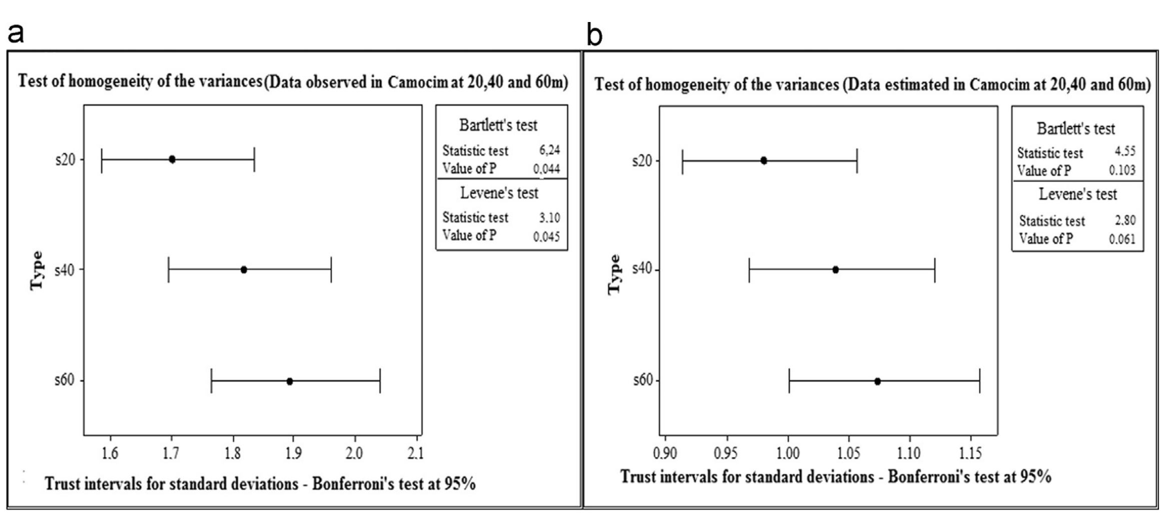

Fig. 28. Test of homogeneity of the variances in region of Camocim: data observed (a) and data estimated (b).

the data series i.e. from August to May, which corresponds to the dry season of the region. This characteristic becomes more evident when increasing the period for the calculation of average wind speeds.

Fig. 17 shows the comparison between data sets obtained by linear regression method and those observed by the AT in Paracuru for daily, 5-day, 10-day, and 15-day averages at the height of 40 m.

By analyzing the aforementioned results, overestimation of data obtained by linear regression if compared with the data observed in the AT generally occurs, which is also found in the calibration period for the same time interval (from August to May).

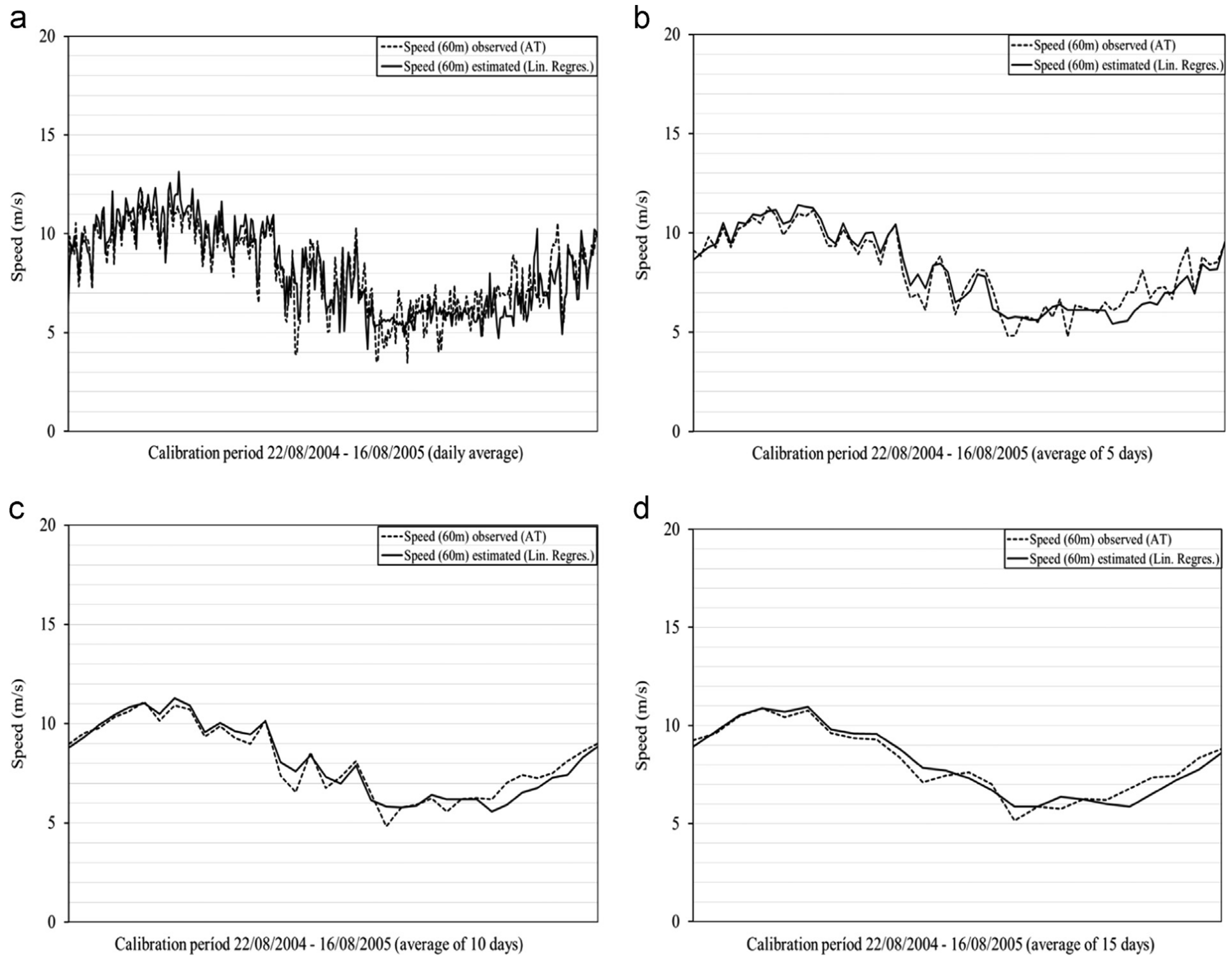

Fig. 29. Comparison between series of data estimated for lineal regression and data observed in AT of Camocim for the height of 60 m in the calibration period, for daily averages (a), of 5 (b), of 10 (c) and of 15 days (d).

Fig. 18 presents the comparisons between the data sets of estimated and observed data by AT of Paracuru within the same aforementioned intervals at the height of 20 m during the validation period.

Thus, one can observe that overestimation of data obtained by regression from those observed by the AT occurs for the same interval of time during the calibration period, as it was mentioned in the previous comparisons.

As in the calibration period, in order to assess the degree of fit between the observed and estimated data in the validation period, parameters RMSE, MAE, and ε are also determined. In addition, the correlation and determination coefficients between the two data sets are calculated. Such results are given in Table 2.

For 20 m, the lowest values of RMSE, MAE, and ε are 0.66 m/s, 0.59 m/s, and 6.3%, respectively, considering the comparison for the 15 day-average. Even at this altitude, the maximum correlation coefficient is 0.99 for 15 days and the maximum average determination coefficient explains 98% of the variability for the observed data.

At 40 m, the lowest values f RMSE, MAE, and ε are 0.66 m/s, 0.56 m/s, and 5.3%, respectively, which are recorded in the comparisons between the 15-day average. The correlation coefficient is maximum (0.98) and the determination coefficient indicates that 96% of the variability of the observed data was captured. For 60 m, the lowest values of the aforementioned errors are 0.72 m/s, 0.62 m/s, and 5.8%, respectively, which correspond to the 15-day average. The observed correlation coefficient is also maximum (0.98) and the maximum determination coefficient explains 96% of the variability for the observed data.

For both calibration and validation periods, the values of RMSE and MAE are lower than the errors found in [3], even when comparing the average wind speeds observed by the AT in Paracuru with the Regional Atmospheric Modeling System (RAMS) at 20 m, 40 m, and 60 m. Additionally, the respective correlation coefficients in both periods are higher than the those obtained in [3].

3.2. *Results for the region of Camocim*

3.2.1. *Characterization of the wind in the region of Camocim*

Figs. 19 and 20 show the daily and monthly profiles of the average wind speed from data collected by the PDC in Barroquinha, respectively.

Regarding the daily profile of the average wind speed, it is observed that higher average wind speeds occur during daylight i.e. between 06:00 and 18:00 local time, with a maximum of 6.2 m/s recorded at around 14:00 h. These results are directly related to the presence of sea breezes, as it is explained in [16,17].

For the monthly profile of the average wind speed, the highest values of average wind speeds occur in the period between August and December. As in the region of Paracuru, this period coincides with the dry season. On the other hand, lower values are observed during the rainy season, specifically between the months from February to May.

Figs. 21 and 22 represent the daily and monthly profiles of the average wind speed in the AT of Camocim, respectively.

For the daily profile, the highest average wind speed is about 12.0 m/s, which occurs at around 17:00 local time.

Considering the monthly profile, the periods where the highest and the lowest average wind speed coincide with the same ones regarding the PDC in Barroquinha i.e. the dry season and the rainy season, respectively. It is also observed that all monthly average values are higher than the entry speed of the wind turbines (2.5 m/s).

Figs. 23 and 24 show the prevailing wind directions at the PDC in Barroquinha and the AT in Camocim. It is possible to notice that the predominant wind direction comes from east, varying between northeast and southwest directions. This fact evidences the influence of trade winds as stated in [16].

3.2.2. Calibration period – region of Camocim

Figs. 25, 26, and 27 show the correlation between the estimated average wind speeds for 60 m, 40 m, and 20 m, respectively, which are obtained from observed data in the surface by the PDC in Barroquinha and also the values observed by the AT in Camocim.

In addition, it is shown that line equations for the linear regression fits the determination coefficients (R^2) obtained for daily averages of 5, 10, and 15 days. The determination coefficient also increases when the calculation period of the average wind speed is extended, as it occurs in the region of Paracuru.

Fig. 25 presents some results for daily wind speed at the height of 60 m (Fig. 25a), where the linear regression model explains approximately 82% of the variability of the observed data. On the other hand, the model explains about 93% of such variability (Fig. 25d).

Fig. 26 shows the results for 40 m, where the model explains 82% of the variability of the observed data for the daily average (Fig. 26a). Besides, the model explains 94% of such variability for the 15-day average (Fig. 26d).

The results for 20 m are shown in Fig. 27. For the daily average (Fig. 27a) and 15-days average (Fig. 27d), the model explains 83% and 95% of the variability of the observed data, respectively.

A homogeneity test for the variance between samples was also carried out. Fig. 28 shows the results by using two methods: Bartlett and Levene. Moreover, a confidence interval for standard deviations exists according to the Bonferroni's test. In the test for the observed values, $P<0.05$ is obtained, thus leading to the rejection of hypothesis H_0 i.e. the variance of samples is not homogenous, but the obtained values ($P=0.045$ and $P=0.044$) are so close to 0.05 that it is reasonable to accept them. One explanation for obtaining such values is based on the fact that the number of samples in the region of Camocim is less than that treated in Paracuru. Besides, some inconsistence is identified in the observed samples and even a lot of missing data is fixed by using the gap filling method.

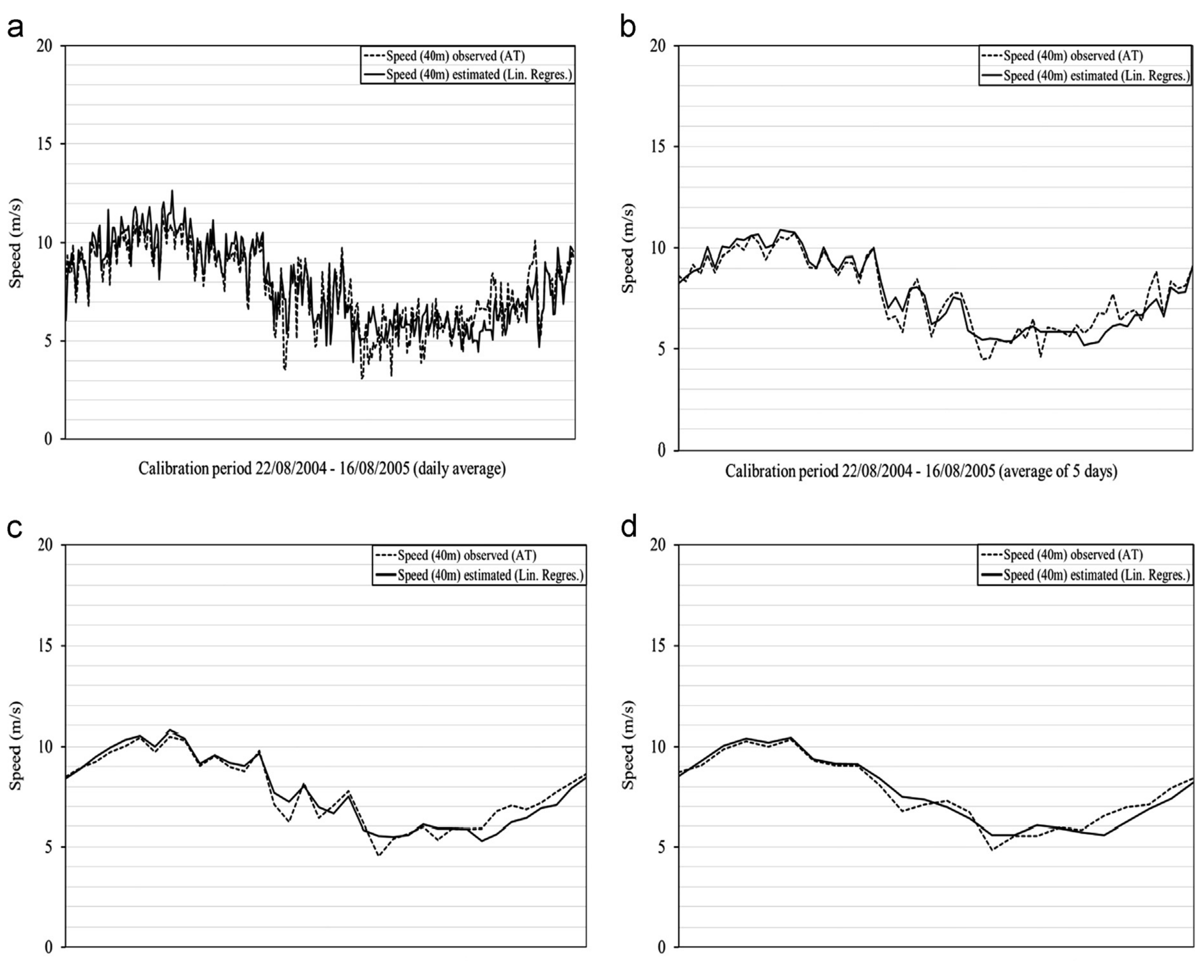

Fig. 30. Comparison between series of data estimated for lineal regression and data observed in AT of Camocim for the height of 40 m in the calibration period, for daily averages (a), of 5 (b), of 10 (c) and of 15 days (d).

Analogously to the region of Paracuru, after obtaining the lines representing the fit of linear regression between the datasets of wind speed, which were previously shown in the calibration period, estimated values are obtained for such straight lines at three heights at the AT in Camocim, as the results are compared with the observed data sets.

Fig. 29 presents the comparison between data sets obtained by linear regression method and those observed by the AT in Camocim for daily averages during 5, 10, and 15 days at 60 m. As observed in the region of Paracuru, the data series obtained by linear regression model in general overestimates the data series observed in the dry season while underestimating the rainy season.

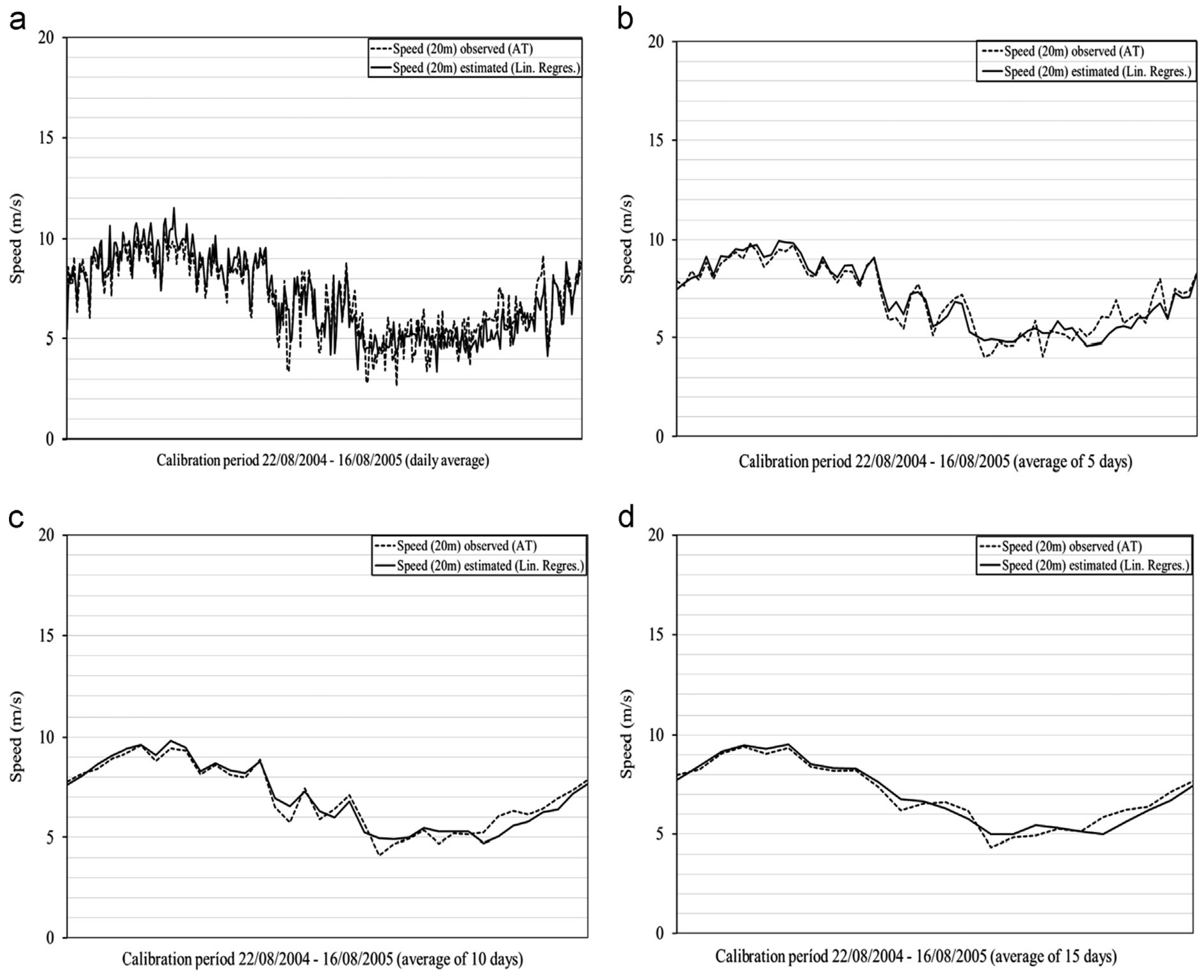

Fig. 31. Comparison between series of data estimated for lineal regression and data observed in AT of Camocim for the height of 20 m in the calibration period, for daily averages (a), of 5 (b), of 10 (c) and of 15 days (d).

Table 3
Values of the errors and correlation and determination coefficients between the data estimated in AT for the equations of the regression line and the data observed in AT in the calibration period for the region of Camocim.

Height (m)	Periods which it was obtained the averages of the speeds of the wind	RMSE (m/s)	MAE (m/s)	ε (%)	r	R^2
20 m	Daily	0.80	0.61	0.56	0.91	0.83
	5 Days	0.50	0.39	0.24	0.95	0.90
	10 Days	0.39	0.31	0.19	0.97	0.94
	15 Days	0.36	0.28	0.06	0.97	0.94
40 m	Daily	0.92	0.69	1.16	0.89	0.79
	5 Days	0.55	0.41	0.04	0.95	0.90
	10 Days	0.44	0.33	0.04	0.97	0.94
	15 Days	0.40	0.30	0.02	0.97	0.94
60 m	Daily	0.94	0.71	0.87	0.90	0.81
	5 Days	0.57	0.44	0.03	0.95	0.90
	10 Days	0.47	0.36	0.16	0.97	0.94
	15 Days	0.42	0.33	0.05	0.97	0.94

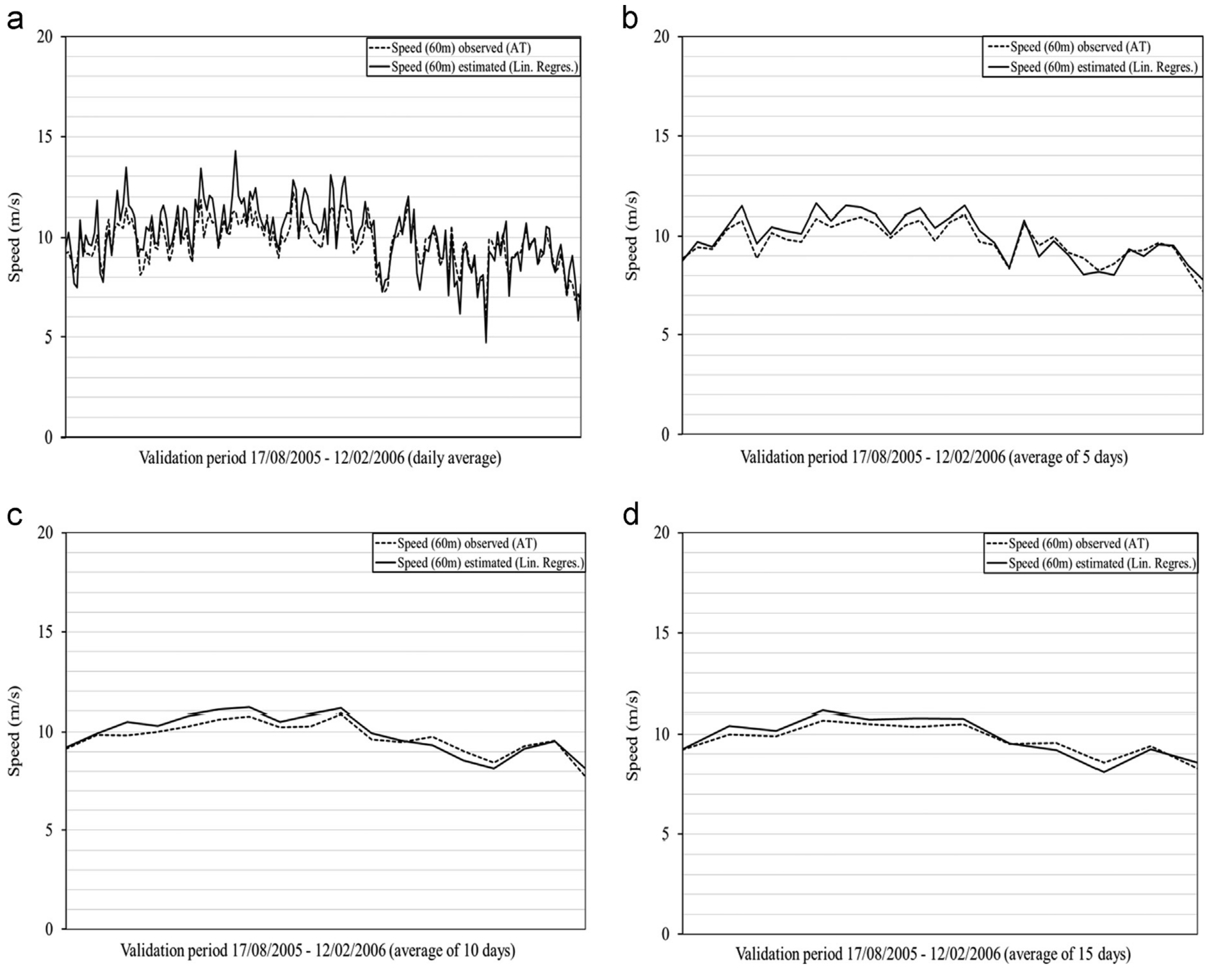

Fig. 32. Comparison between series of data estimated for lineal regression and data observed in AT of Camocim for the height of 60 m in the validation period, for daily averages (a), of 5 (b), of 10 (c) and of 15 days (d).

Fig. 30 corresponds to the comparison between data sets obtained by linear regression method and those observed in AT of Camocim for daily averages of 5, 10, and 15 days at the height of 40 m. Once again, the results evidence periods of underestimation and overestimation similar to the aforementioned ones.

Fig. 31 shows the comparison between the data series estimated by linear regression method and those observed in AT Camocim for daily averages of 5, 10 and 15 days at the height of 20 m. In this analysis, overestimation of the observed values in dry season and underestimation in the rainy season also occur as in the previous comparisons for 60 m and 40 m.

The quantification of errors is performed by using the same methods in the region of Paracuru. Table 3 shows the calibration period, the quantification results, and also the correlation and determination coefficients.

The error values are lower and the determination coefficients are higher if compared with the errors in the calibration period for the region of Paracuru.

For the estimation at 20 m high, the lowest values of RMSE, MAE, and ε are 0.36 m/s, 0.28 m/s, and 0.06%, respectively, which are recorded for the comparison between sets of estimated and observed data during 15 days. Even at such altitude, the maximum correlation coefficient is 0.97 for the comparison between observed and estimated average values during 10 and 15 days, while the maximum determination coefficient explains 94% of the variability for the observed data.

At 40 m, the lowest values of RMSE, MAE, and ε are 0.40 m/s, 0.30 m/s, and 0.02%, respectively, which are recorded for the comparison between sets of estimated average values and observed ones during 15 days. The maximum correlation coefficient is 0.97 and the maximum determination coefficient explains 94% of the variability of the observed data.

At 60 m, the lowest values of the aforementioned errors are 0.42 m/s, 0.33 m/s, and 0.03%, respectively. The first two parameters were recorded for the calculation period of 15 days and last one was obtained during 10 days. At 60 m the correlation coefficient is also maximum (0.97) and the maximum determination coefficient explains 94% of the variability of the observed data.

For the region of Camocim, the Kolmogorov–Smirnov's test was also performed (2003), which has shown significance of 95%.

3.2.3. Validation period – region of Camocim

Fig. 32 presents the comparison between data sets obtained by linear regression method and those observed in AT of Camocim during the validation period for daily averages of 5, 10 and, 15 days at 60 m. It also shows that the data series obtained by linear regression model generally overestimates the observed one regarding the AT, what becomes more evident as the period for the calculation of average wind speeds is extended.

These results are consistent with those presented in the calibration period, which also present overestimation within the same time intervals, corresponding to dry season of the region.

Fig. 33 shows the comparison between the data sets obtained by linear regression method and those observed by the AT in

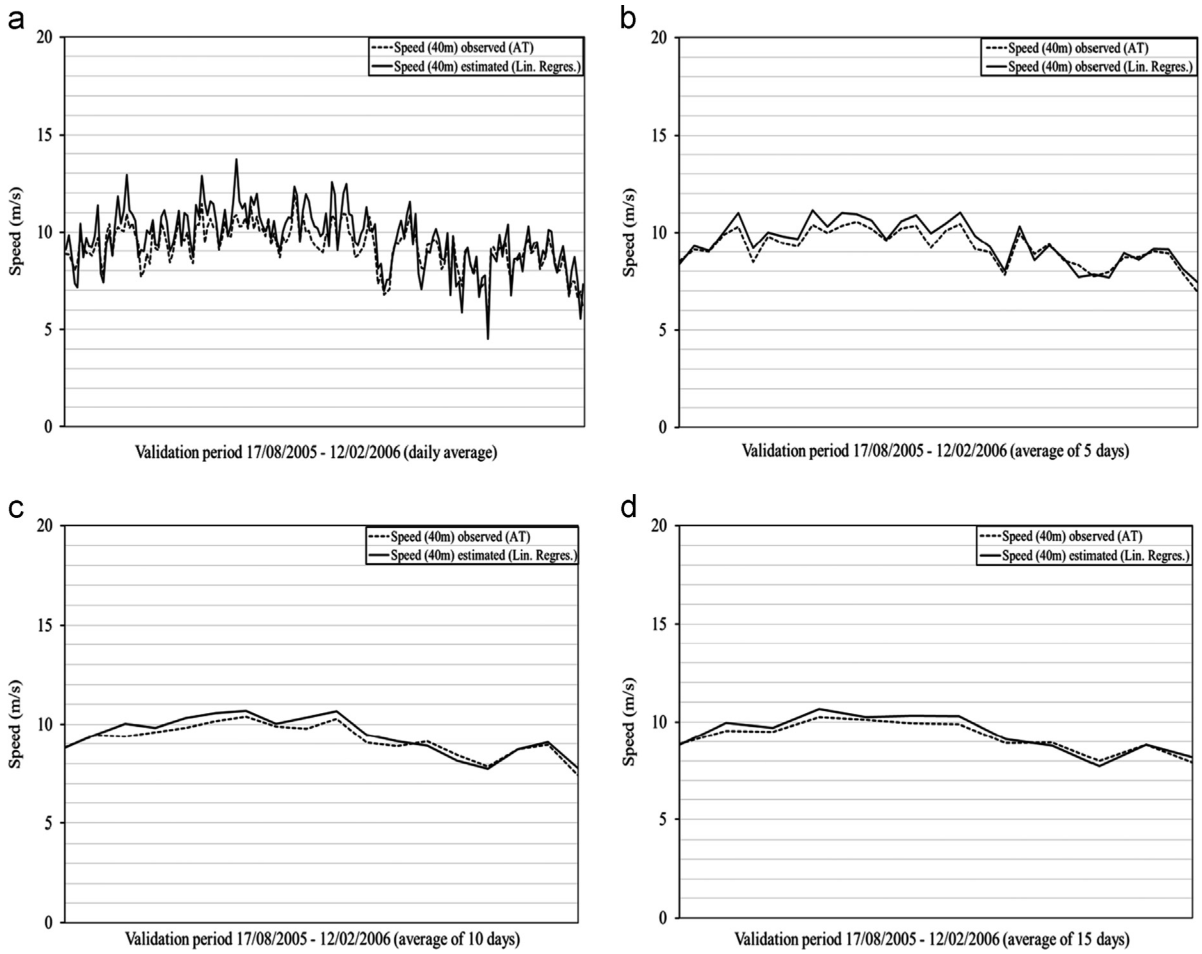

Fig. 33. Comparison between series of data estimated for lineal regression and data observed in AT of Camocim for the height of 40 m in the validation period, for daily averages (a), of 5 (b), of 10 (c) and of 15 days (d).

Camocim for daily averages of 5, 10 and 15 days at the height of 40 m. By analyzing the results, there is also an indication of overestimation and underestimation of the observed data series within the same time intervals observed for the calibration period.

Fig. 34 shows the comparison between the data sets obtained by linear regression method and those observed by the AT in Camocim at the height of 20 m. In general, there is good agreement between the observed and estimated data series, while overestimation and underestimation exist within the same time intervals during the calibration period.

In addition to the validation period, the degree of fit between the observed and estimated data is assessed by quantifying the same errors that were considered for the region of Paracuru, as well as the correlation and determination coefficients between the two data sets. The relevant results are then shown in Table 4.

Lower values of RMSE, MAE, and ε occur at 20 m high, which are 0.24 m/s, 0.21 m/s, and 1.0%, respectively. The parameters are determined for comparisons between series of average estimated and observed data during 15 days. The maximum correlation coefficient is 0.98 and the maximum determination coefficient explains 96% of the variability for the observed data.

At 40 m, the lowest values of RMSE, MAE, and ε are 0.28 m/s, 0.25 m/s, and 1.8%, respectively, for the comparison between the series of estimated and observed data during 15 days. The maximum correlation coefficient is 0.98 and the maximum determination coefficient is 0.96, thus explaining about 96% of the variability for the observed data.

For 60 m, the lowest errors are 0.32 m/s, 0.29 m/s, and 1.2%, respectively for the comparison between the series of estimated and observed data during 15 days. The maximum correlation coefficient is 0.97, while the maximum determination coefficient explains 94% of the variability for the observed data.

Once again, for both the calibration period and the validation period, the RMSE and MAE are lower than the values found in [3], when comparing the average wind speeds observed by the AT in Camocim with the values simulated with the RAMS at 20 m, 40 m, and 60 m. However, the presented correlation coefficients coincide with the values found in [3]. It is worth to mention that the correlation coefficients and their results obtained via regression line and adjusted during both calibration and validation periods for the 10-day and 15-day average must be interpreted carefully since PDC-related data presents some inconsistences during the studied period. Besides, equipment has been installed recently, while the samples the aforementioned period are too few to fit the regression line and provide statistically reliable results.

4. Conclusion

The linear regression model on which this work is based has shown satisfactory performance within both the calibration period and the validation period due to the high rate of agreement between the estimated data sets and also to the observed data and their correlation coefficients, with consequent low values of the quantified errors.

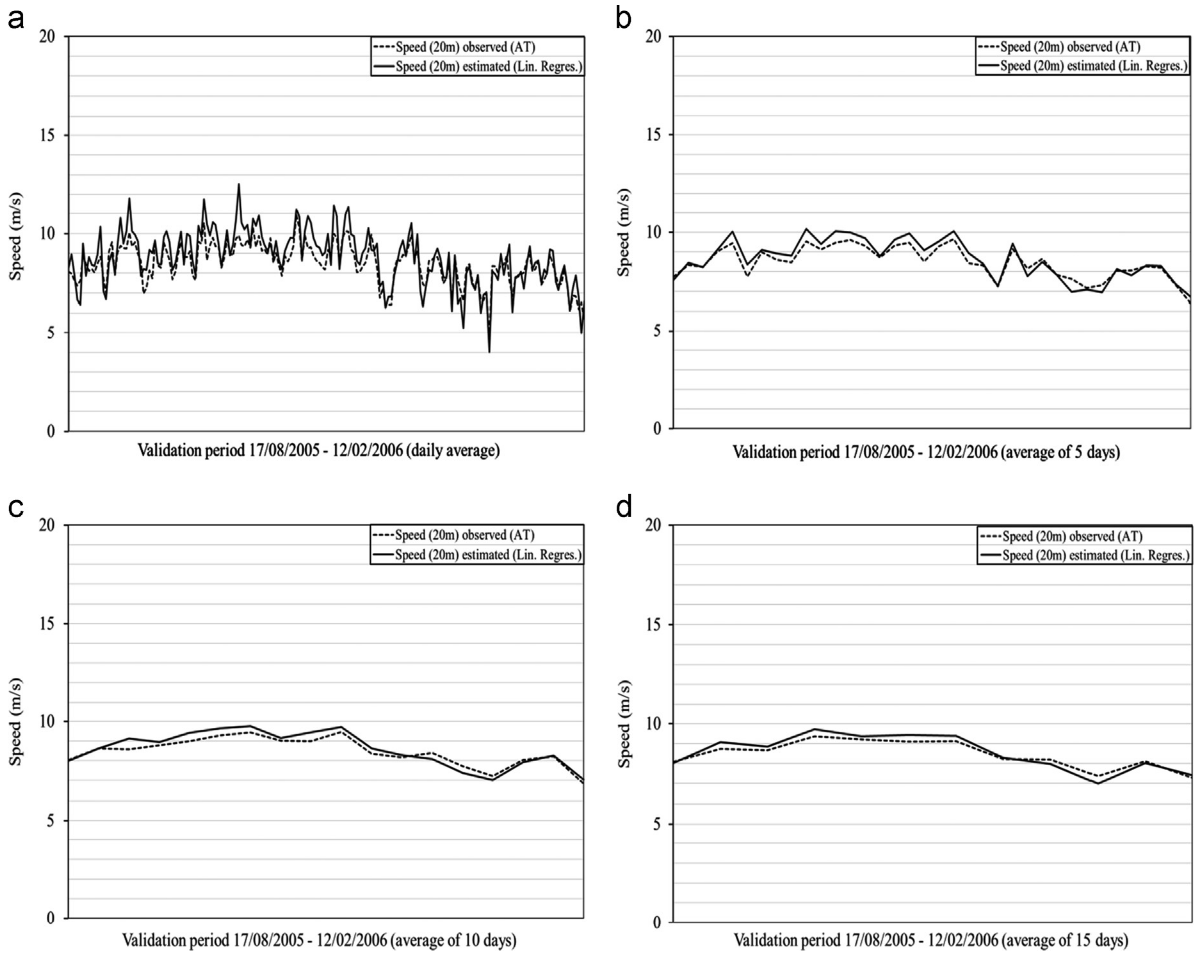

Fig. 34. Comparison between series of data estimated for lineal regression and data observed in AT of Camocim for the height of 20 m in the validation period, for daily averages (a), of 5 (b), of 10 (c) and of 15 days (d).

Table 4
Values of the errors and correlation and determination coefficients between the data estimated in AT for the equations of the regression line and the data observed in AT in the validation period for the region of Camocim.

Height (m)	Periods which it was obtained the averages of the speeds of the wind	RMSE (m/s)	MAE (m/s)	ε (%)	r	R^2
20 m	Daily	0.72	0.58	3.7	0.90	0.81
	5 Days	0.34	0.28	1.9	0.97	0.94
	10 Days	0.28	0.24	1.5	0.97	0.94
	15 Days	0.24	0.21	1.0	0.98	0.96
40 m	Daily	0.82	0.66	4.6	0.90	0.81
	5 Days	0.41	0.35	2.7	0.96	0.92
	10 Days	0.33	0.28	2.2	0.97	0.94
	15 Days	0.28	0.25	1.8	0.98	0.96
60 m	Daily	0.84	0.68	3.7	0.89	0.79
	5 Days	0.44	0.37	2.1	0.95	0.90
	10 Days	0.37	0.32	1.7	0.95	0.90
	15 Days	0.32	0.29	1.2	0.97	0.94

Although the model has been used in the coastal areas of Ceará, it can be further applied to the analysis of data from PDCs located in other regions.

Based on the obtained results, it has been confirmed that the regions of Paracuru and Camocim can be characterized as potential sites adequate for wind energy generation. It can be stated that the estimation of wind resources in altitudes from data observed at the surface is possible by using the methodology proposed in this work, which consists in a technically and financially viable methodology.

Future work includes the use of multiple linear regression i.e. a single dependent variable, but with two or more independent variables in order to improve the estimation capacity. Another possibility lies in quantifying the wind energy potential from average wind speeds observed by the AT and the values estimated by linear regression from the data surface.

References

[1] Dobriansky P. Clean energy for the future. J Econ Perspect 2006;11(2):4–6.
[2] Oliveira JL. Influence of general circulation and the interannual variability of the wind energy potential in Northeast Brazil. Thesis [Master of Science in Applied Physics], University of Ceará; 2007.
[3] De Maria Santiago, Costa PHS, Alexandre A., Sombra Sergio S. Simulation of high resolution winds on the coast of Ceará. In: Brazilian congress of meteorology, proceeding of XIV Congress Bras. Meteorology. SEINFRA. Anemometer data (wind measurement) – Paracuru and Camocim. Available at: 〈http://www.seinfra.ce.gov.br/index.php/downloads/category/3-energia〉; 2006 [accessed 25.05.09].
[4] Kousky VE. Front influence in Northeast Brazil. Mon Weather Rev 1979;107:1140–1153.
[5] Andreoli RV, Kayano MT, Guedes RL, Oyama MD, Alves MAS. The influence of sea surface temperature of the Pacific and Atlantic Oceans on precipitation variability in Fortaleza. Mag Braz Meteorol 2004;19(2):113–22.
[6] Xavier TMS, Xavier AF, Dias Da Silva PL, Da Silva Dias MAF. The intertropical convergence zone – ITCZ and its relations with rain in Ceará (1964–98). Mag Braz Meteorol 2000;15(1):27–43.
[7] Rao VB, Lima MC, Franchito S. Seasonal and interannual rainfall in Northeast of Brazil. J Clim 1993;6:1754–63.
[8] Hirata MH, Moura AD, Bachmann J. The study of the influence of extra-tropical latitudes systems on the climatic variability on northeast Brazil. Mag Braz Meteorol 1986;1(1):11–7.
[9] Stull RB. An introduction to boundary layer meteorology. Dordrecht: Kluwer Academic Publishers; 1988.
[10] Roballo ST, Fisch G. Atmospheric flow at the Alcantara Launch Center (ALC): Part I: observational aspects. Mag Braz Meteorol 2008;23(4):510–9.
[11] Alcantara CR, Souza EP. A thermodynamic theory for breezes: tests using numerical simulations. Mag Braz Meteorol 2008;23(1):1–11.
[12] Teixeira RFB. The phenomenon of breeze and its relation to rain on Fortaleza-CE. Mag Braz Meteorol 2008;23(3):282–91.
[13] Carvalho P. Wind generation. Fortaleza - Ceará- Brazil: Fortaleza: Imprensa Universitária; 2003 (146 p.).
[14] Castro RMG. Renewable energy and decentralized production: introduction to wind energy. 3rd ed. Lisboa: Technical University of Lisboa; 2007.
[15] Carvalheiro, LCO, et al. Validation of sea surface temperature from satellite measurements. Available at: 〈www.ame-web.org/JORNADAS/6B_Carvalheiro.pdf〉 [accessed15.09.09].
[16] Barreto AB, Aragon MRS, Braga CC. Study of diurnal cycle of surface winds in the Northeast of Brazil. In: Brazilian congress of meteorology, proceedings of XII Brazilian congress of meteorology. Foz do Iguaçu; 2002.
[17] Molion LCB, Bernard SO. Variability of the wind to the city of Maceió-AL. In: Brazilian congress of meteorology, proceedings of XI Brazilian congress of meteorology. Rio de Janeiro; 2000.
[18] Sacramento, EM, et al. Preliminary study on the potential of wind from the mountain regions of Ceará. In: Brazilian congress of meteorology, proceedings of XIV Brazilian congress of meteorology. Florianópolis; 2006.
[19] Devore DL. Probability and statistics for engineering and science. São Paulo - Brazil: Thompson Learning Thomson Learning; 2006.
[20] Pinto MO. Fundamentals of wind energy. Rio de Janeiro: LTC; 2013.
[21] Kolmogorov-Smirnov. Available at: 〈www.physics.csbsju.edu/stats/KS-test.html〉 [accessed 15.10.09].