Meteorol. Appl. 13, 29–41 (2006) doi:10.1017/S1350482705001982

Comparative analysis of estimators for wind direction standard deviation

P. S. Farrugia & A. Micallef
Department of Physics, University of Malta, Msida MSD 06, Malta
Email: pierresandrefarrugia@yahoo.co.uk; alfred.micallef@um.edu.mt

Wind direction is a circular variable. This makes the algorithms used to find its standard deviation different from that of the linear variables. In particular, the requirement for storing all the data points before the standard deviation can be computed limits the storage capacity and puts great strain on remote data acquisition systems. Various algorithms have therefore been developed to estimate the standard deviation in order to reduce the number of terms stored. The following work consists of a comparative analysis of such estimators together with the parameters used. It emerges that some of the assumptions adopted to produce the equations being analysed do not hold in practice, even though this does not affect significantly the performance of the estimators that depend on them. On the other hand, the parameter that has the best trend with the algorithm adopted is the magnitude of the vector to the centre of gravity of the system. However, such a result gives rise to some concerns since it does not account for the 'vectorial' nature of the angle being treated.

Keywords: wind direction, estimator, angular standard deviation, circular variables

Received January 2004, revised September 2005

1. Introduction

Wind is a vector quantity having both a magnitude and a three-dimensional direction. This would make wind a spherical variable. However, usually only the horizontal component is considered. Thus, wind is mainly treated as a circular variable with an associated magnitude.

Circular variables are different from the more commonly treated linear variables. They are periodic, the zero is arbitrary and the notion of high and low values cannot be applied. This results in the need for a different statistical treatment. In the treatment of wind direction there are various algorithms that are employed to find the standard deviation (Yamartino 1984; Essenwanger 1985). The choice depends on the experimenter.

The problematic feature of such algorithms is that they require the storage of all data points to be computed. This puts a burden on the storage capacity which is limited in remote data acquisition systems. Thus there is a need to estimate such algorithms using a minimum number of representative terms.

Various estimators have been developed to fulfil this function. In general, most of them seem to show substantial agreement, especially at low values of dispersion. However, for larger angular variation, there is a progressive increase in the range of values of the algorithm that corresponds to a single value of the parameter being analysed. The reasons for this are various and can stem from the techniques used, such as limit in the approximations employed. As for theoretical derivations, there is always a very important underlying difficulty. Wind variation depends on a substantial number of variables that include solar irradiance, latitude, terrain and mechanical interactions, convection currents and the Coriolis force. Thus, samples taken at different locations and/or times seem to come from different populations. This makes it difficult to generalise. It is therefore important to test the assumptions made using real data, something that is frequently not done.

An analysis of the procedure used reveals there is no general agreement as to whether the biased or unbiased sample variance should be adopted. In linear statistics the latter is employed. However, for wind data, this seems to be an arbitrary matter, with some choosing to use the former (Yamartino 1984), and others adopting the latter (Verrall & Williams 1982; Weber 1992). Under such circumstances, we decided to adopt the definition of the unbiased sample variance, with the proviso that the equations were presented in such a way so as to make the conversion relatively easy.

Another relevant point is that the algorithm used to test the estimators developed is not often mentioned, thereby making it difficult, if not impossible, to compare and contrast their performance. This is particularly relevant for the case of data fitting, where the constants derived would depend to some extent on the algorithm being adopted.

2. Summary of the parameters used

The estimators used to find the standard deviation of wind direction are usually based on parameters that can be computed in a single pass. Thus it is necessary to make a brief summary of the quantities that will be used. These can be broadly divided into two. One set of parameters is related to the mean of the various quantities measured while the other refers to their variance.

Starting with the former, a statistic that is frequently used as a measure of the mean for circular variables is the mean angle of the sample (θ_a). This is found by mapping each angle, θ_i, to the polar coordinates, (θ_i,1), and then finding the angle that the vector to centre of gravity of the system (**R**) makes with the positive x-axis. Since in meteorology the angle is measured clockwise from the north, it will be convenient to align the x-axis with this direction and the y-axis with east. In such a coordinate system **R** will have Cartesian components

$$\overline{C} = n^{-1} \sum_{i=1}^{n} \cos(\theta_i) \tag{1}$$

$$\text{and } \overline{S} = n^{-1} \sum_{i=1}^{n} \sin(\theta_i), \tag{2}$$

along the x- and y-axes respectively. Thus θ_a will be given by

$$\theta_a = \arctan(\overline{S}/\overline{C}), \tag{3}$$

taking into account the quadrant in which the $\overline{C}$ and $\overline{S}$ reside. The magnitude of the vector to the centre of gravity (R) would then be given by

$$R = (\overline{S}^2 + \overline{C}^2)^{1/2}, \tag{4}$$

and its possible values go from zero to one. Since the wind has an associated magnitude, it is possible to define a weighted mean angle by taking the angle that the mean wind vector ($\overline{\mathbf{V}}$) makes with the positive x-axis. Now $\overline{\mathbf{V}}$ has Cartesian components along the x- and y-axis given respectively by

$$\overline{V}_x = n^{-1} \sum_{i=1}^{n} V_i \cos(\theta_i) \tag{5}$$

$$\text{and } \overline{V}_y = n^{-1} \sum_{i=1}^{n} V_i \sin(\theta_i). \tag{6}$$

Hence the weighted mean angle of the sample, θ_v, would be given by

$$\theta_V = \arctan(\overline{V}_y / \overline{V}_x), \tag{7}$$

where again care must be taken of the quadrant in which the $\overline{V}_x$ and $\overline{V}_y$ are. The magnitude of the mean vector speed ($|\overline{\mathbf{V}}|$) is then given by

$$|\overline{\mathbf{V}}| = \left(\overline{V}_x^2 + \overline{V}_y^2\right)^{1/2}. \tag{8}$$

This is different from the mean wind speed ($\overline{V}$), which is defined as

$$\overline{V} = n^{-1} \sum_{i=1}^{n} V_i. \tag{9}$$

Their ratio, which is called the persistence (P), gives a measure of the tendency of the wind to blow in one direction, and is computed using

$$P = |\overline{\mathbf{V}}|/\overline{V}. \tag{10}$$

Its value ranges from zero to one.

Considering now the parameters that are based on the variance of the quantities measured, these include:

the variance of the cosine of the angles:

$$s_C^2 = [n/(n-1)] \left[n^{-1} \sum_{i=1}^{n} \cos^2(\theta_i) - \overline{C}^2 \right] \tag{11}$$

the variance of the sine of the angles:

$$s_S^2 = [n/(n-1)] \left[n^{-1} \sum_{i=1}^{n} \sin^2(\theta_i) - \overline{S}^2 \right] \tag{12}$$

the x-velocity variance:

$$s_{Vx}^2 = [n/(n-1)] \left[n^{-1} \sum_{i=1}^{n} V_i^2 \cos^2(\theta_i) - \overline{V}_x^2 \right] \tag{13}$$

the y-velocity variance:

$$s_{Vy}^2 = [n/(n-1)] \left[n^{-1} \sum_{i=1}^{n} V_i^2 \sin^2(\theta_i) - \overline{V}_y^2 \right] \tag{14}$$

and the x-y-velocity covariance:

$$s_{Vxy}^2 = [n/(n-1)] \left[n^{-1} \sum_{i=1}^{n} V_i^2 \sin(\theta_i) \cos(\theta_i) - \overline{V}_x \overline{V}_y \right]. \tag{15}$$

Rotating the coordinate system by θ_v, and mapping the x-axis and the y-axis to the u-axis and v-axis gives:

the along-wind variance:

$$s_{Vu}^2 = [n/(n-1)] \left[n^{-1} \sum_{i=1}^{n} V_i^2 \cos^2(\theta_i - \theta_V) - \left|\overline{\mathbf{V}}\right|^2 \right]. \tag{16}$$

and the cross-wind variance:

$$s_{Vv}^2 = (n-1)^{-1} \sum_{i=1}^{n} V_i^2 \sin^2(\theta_i - \theta_V). \tag{17}$$

Note that the above quantities refer to the descriptive statistics for the sample. When referring to the population parameters, the symbol σ^2 will be used for the variance, μ for the mean and v for the wind speed.

3. Description of the models

The algorithm used for the comparative analysis was the unbiased form of that suggested by Yamartino (1984), given by

$$s_{Y0}^2 = [n/(n-1)]\left[n^{-1}\sum_{i=1}^{n}\Delta(\theta_a)_i^2 - \left(n^{-1}\sum_{i=1}^{n}\Delta(\theta_a)_i\right)^2\right], \tag{18}$$

where $\Delta(\phi)_i$ is the minimum angular distance between θ_i and ϕ that can be viewed to be

$$\begin{aligned} \Delta(\phi)_i &= \theta_i - \phi \quad \text{such that } |\Delta(\phi)_i| \le \pi, \\ \text{with } \Delta(\phi)_i &> 0 \quad \text{if } \phi \le \theta_i < \phi + \pi, \\ \Delta(\phi)_i &< 0 \quad \text{if } \phi - \pi < \theta_i < \phi, \\ \text{and } \Delta(\phi)_i &= \theta_i - \phi \quad \text{if } |\Delta(\phi)_i| = \pi. \end{aligned} \tag{19}$$

By considering his algorithm and using trigonometric relations and small angle approximations, Yamartino (1984) established that s_{Y0} could be approximated by

Yamartino 1 (Y1):

$$s_{Y1} = [n/(n-1)]^{1/2}(1 - R^2)^{1/2}, \tag{20}$$

where Yamartino 1 is the name given to the estimator and Y1 its abridged form. He found that this expression significantly under predicts s_{Y0} above $1/3\,\pi^c$ and so he proposed the following as a possible correction,

Yamartino 2 (Y2):

$$s_{Y2} = [n/(n-1)]^{1/2}\arcsin\left[(1 - R^2)^{1/2}\right]. \tag{21}$$

However, Yamartino showed that this estimator still diverged for values of s_{Y0} above $1/2\,\pi^c$ and so he tried to find a best-fitting function for the data. To derive this he first considered the maximum value that the standard deviation of wind direction could take. He assumed this to be the standard deviation of a uniformly distributed wind direction that is given by

$$\sigma_\theta = \left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\theta^2 d\theta\right)^{1/2} = \pi/3^{1/2}. \tag{22}$$

He then obtained a best-fitting function using this as the limiting maximum value:

Yamartino 3 (Y3):

$$\begin{aligned} s_{Y3} = [n/(n-1)]^{1/2}&\arcsin\left[(1 - R^2)^{1/2}\right] \\ &\times\left[1 + (2/3^{1/2} - 1)(1 - R^2)^{2/3}\right]. \end{aligned} \tag{23}$$

It should be mentioned that the limiting value for the standard deviation given by Equation 22 is not the maximum value that Equation 18 can give. An adequate counter example is provided by the situation where there are $i+1$ angles with value 0^c and three sets of i angles with values $1/2\,\pi^c$, π^c and $1\,1/2\,\pi^c$ (where the extra angle is need to keep θ_a well defined). In this case, taking the limit as i goes to infinity, gives $s_{Y0}^2 = 3/8\,\pi^{2c}$.

Yamartino also related R to the persistence. By assuming that the wind speed and direction are independent, and for large n, he showed that

$$R \approx P. \tag{24}$$

Using Equation 24, in Equations 20 and 23, he derived the following two equations:

Yamartino P 1 (YP1):

$$s_{YP1} = [n/(n-1)]^{1/2}(1 - P^2)^{1/2}. \tag{25}$$

Yamartino P 2 (YP2):

$$\begin{aligned} s_{YP2} = [n/(n-1)]^{1/2}&\arcsin\left[(1 - P^2)^{1/2}\right] \\ &\times\left[1 + (2/3^{1/2} - 1)(1 - P^2)^{2/3}\right]. \end{aligned} \tag{26}$$

Yamartino did not actually present Equations 25 and 26 as estimators for s_{Y0}. However, others considered them in this way (Mori 1986; Weber 1992).

Mardia (1972) derived a theoretical estimate of the standard deviation of a circular variable by considering the wrapped normal distribution. His estimator was

Mardia 1:

$$s_{\text{Mardia 1}} = [-2\ln(R)]^{1/2}. \tag{27}$$

The problem with this predictor is that it diverges to infinity as R goes to 0 resulting in substantial over-prediction. However, for s_{Y0}, there is a definite upper bound constituted by π^c. This discrepancy arises from a different concept of the standard deviation that, in the case of the wrapped normal distribution, is allowed to go to infinity. In order to improve the agreement of Mardia 1 with s_{Y0}, it was assumed that $\exp(-s_\theta^2)$ was some function f of R. Given the range of R, the following boundary conditions were applied:

$$f(1) = 1 \tag{28}$$

$$\text{and } f(0) = \exp(-\pi^2/3), \tag{29}$$

where the maximum standard deviation of the wind direction was estimated to be that of the uniform distribution.

Since R is relatively small, it was concluded that f could be approximated by a linear equation. In this case two intuitively appealing solutions would be

Predictor 1 (P1):

$$s_{P1} = \left[-\ln\left\{\left[1 - \exp\left(-\tfrac{1}{3}\pi^2\right)\right]R^2 + \exp\left(-\tfrac{1}{3}\pi^2\right)\right\}\right]^{1/2}. \tag{30}$$

Predictor 2 (P2):

$$s_{\mathrm{P2}} = \left[-2\ln\left\{\left[1-\exp\left(-\tfrac{1}{6}\pi^2\right)\right]R + \exp\left(-\tfrac{1}{6}\pi^2\right)\right\}\right]^{1/2}. \quad (31)$$

However, these are not the only predictors that can be obtained with this procedure. In fact the boundary conditions set by Equations 28 and 29 will be met by any polynomial $a_0 + a_1 R + \ldots + a_k R^k$ satisfying

$$a_0 = \exp(-\pi^2/3) \quad (32)$$

$$\text{and} \sum_{i=1}^{k} a_i = 1 - \exp(-\pi^2/3), \quad (33)$$

of which there are an infinite number.

Ackermann (1983) derived a theoretical estimate for the standard deviation of wind direction by considering the wind vector. From Equation 7, he inferred that θ_{v} is a function of two variables, $\overline{V}_x$ and $\overline{V}_y$, and using the relation that exists between the standard deviation of a quantity and that of the variables on which it depends (Baird 1962), he derived

$$s_{\mathrm{Ack}} = \left[\overline{V}_y^2 s_{Vx}^2 + \overline{V}_x^2 s_{Vy}^2 - 2\overline{V}_x\overline{V}_y s_{Vxy}^2\right]^{1/2} / |\overline{\mathbf{V}}|^2. \quad (34)$$

Yamartino (1984) showed that this expression could be rewritten more compactly as,

Ackermann (Ack):

$$s_{\mathrm{Ack}} = s_{Vv}/|\overline{\mathbf{V}}|. \quad (35)$$

Another theoretical estimator of the standard deviation of wind direction was derived by Verrall & Williams (1982) by assuming that the angles were normally distributed. They obtained two vectors, namely:

$$\mathbf{v}_1 = [\overline{S} - s_{\mathrm{S}}, \overline{C} + s_{\mathrm{C}}] \quad (36)$$

$$\text{and } \mathbf{v}_2 = [\overline{S} + s_{\mathrm{S}}, \overline{C} - s_{\mathrm{C}}], \quad (37)$$

which they assumed to be symmetrically distributed about the mean angle, at an angle equal to the standard deviation. So the angle between them would correspond to twice the standard deviation of wind direction, which can thus be found using the dot product as

Verrall & Williams (VW):

$$s_{\mathrm{VW}} = 0.5\arccos[(\mathbf{v}_1.\mathbf{v}_2)/(|\mathbf{v}_1|.|\mathbf{v}_2|)]. \quad (38)$$

Yamartino (1984) examined the performance of their equation and concluded that there was an under prediction at high values of s_{Y0} and deduced that this could be reduced by the following correction to the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$:

$$\mathbf{v}_1 = [\overline{S} - \gamma\, s_{\mathrm{S}}, \overline{C} + \gamma\, s_{\mathrm{C}}] \quad (39)$$

$$\text{and } \mathbf{v}_2 = [\overline{S} + \gamma\, s_{\mathrm{S}}, \overline{C} - \gamma\, s_{\mathrm{C}}], \quad (40)$$

so that Equation 38 would then become as follows:

Corrected Verrall & Williams (CVW):

$$s_{\mathrm{CVW}} = (2\gamma)^{-1}\arccos[(\mathbf{v}_1.\mathbf{v}_2)/(|\mathbf{v}_1|.|\mathbf{v}_2|)], \quad (41)$$

where γ is equal to $3^{1/2}/2$.

Mori (1986) investigated the relation between the persistence and the standard deviation of wind direction, and using theoretical relations derived the following estimator:

Mori 1:

$$s_{\mathrm{Mori\,1}} = \arctan\left\{[2(1-P)/P]^{1/2}\right\}. \quad (42)$$

He then turned to Equation 24, which he used in Equation 27 to obtain

Mori 2:

$$s_{\mathrm{Mori\,2}} = [-2\ln(P)]^{1/2}. \quad (43)$$

Considering what Mori did to obtain Equation 43, it can be deduced that it should be possible to interchange P and R in any equation. Thus Equation 42 was rewritten in terms of R giving

Mori 1 R:

$$s_{\mathrm{Mori\,1\,R}} = \arctan\left\{[2(1-R)/R]^{1/2}\right\}. \quad (44)$$

In addition, Predictor 1 and Predictor 2 were converted to their P equivalent, giving

Predictor 3 (P3):

$$s_{\mathrm{P3}} = \left[-\ln\left\{\left[1-\exp\left(-\tfrac{1}{3}\pi^2\right)\right]P^2 + \exp\left(-\tfrac{1}{3}\pi^2\right)\right\}\right]^{1/2}. \quad (45)$$

Predictor 4 (P4):

$$s_{\mathrm{P4}} = \left[-2\ln\left\{\left[1-\exp\left(-\tfrac{1}{6}\pi^2\right)\right]P + \exp\left(-\tfrac{1}{6}\pi^2\right)\right\}\right]^{1/2}. \quad (46)$$

Weber (1992) obtained the following as the best-fitting function between the standard deviation of wind direction and persistence:

Weber (WB):

$$s_{\mathrm{WB}} = 1.8457(1-P)^{0.5337}. \quad (47)$$

The corresponding R equivalent of this equation is

Weber R (WBR):

$$s_{\mathrm{WBR}} = 1.8457\,(1-R)^{0.5337}. \quad (48)$$

Leung & Liu (1995) took the expression proposed by

Weber (1992) and found that the following equation fitted their data better:

Leung & Liu (LL):

$$s_{\mathrm{LL}} = 1.693(1 - P)^{0.46}. \tag{49}$$

This gives the R equivalent

Leung & Liu R (LLR):

$$s_{\mathrm{LLR}} = 1.693(1 - R)^{0.46}. \tag{50}$$

The last existing models for the standard deviation of wind direction to be considered were derived from the Isotropic Gaussian model. Here it is assumed that the components of the wind along the x- and y-directions are described by a bivariate normal distribution, with each variable having the same standard deviation (i.e. $2\sigma_{Vx}^2 = 2\sigma_{Vy}^2 = \overline{\sigma}_V^2$). In addition, the axes are rotated in such a way as to make the correlation coefficient equal to zero, i.e. the wind components become independent. The resulting probability density function is as follows:

$$P(v_x, v_y) = \left(2\pi\overline{\sigma}_V^2\right)^{-1} \exp\left\{-[(v_x - \mu_{Vx})^2 + (v_y - \mu_{Vy})^2]/\left(2\overline{\sigma}_V^2\right)\right\}. \tag{51}$$

Changing the variables v_x and v_y to v and θ and integrating θ^2 over all possible values would give the variance. This was done using analytical and numerical methods by Castans & Barquero (1994) giving the following estimator,

Castans & Barquero (CB):

$$s_{\mathrm{CB}} = \left(\pi/3^{1/2}\right) \exp\left[-1.242\left\{\left[(|\overline{\mathbf{V}}|/\overline{s}_V)^{2.35} + \left(\pi^{1/2}/2\right)^{2.35}\right]^{1/2.35} - \left(\pi^{1/2}/2\right)\right\}^{0.58}\right], \tag{52}$$

where, for computational purposes,

$$\overline{s}_V = \left(s_{Vx}^2 + s_{Vy}^2\right)^{1/2}. \tag{53}$$

Ibarra (1995) used a relation between the mean and vector mean wind speed obtained by a similar procedure, which he substituted in Equation 52 to obtain,

Ibarra 1 (IB1):

$$s_{\mathrm{IB1}} = \left(\pi/3^{1/2}\right) \exp\left[-1.242\left\{(\overline{V}/\overline{s}_V) - \left(\pi^{1/2}/2\right)\right\}^{0.58}\right]. \tag{54}$$

Then he did the same thing with the persistence, obtaining the following estimator:

Ibarra 2 (IB2):

$$s_{\mathrm{IB2}} = \left(\pi/3^{1/2}\right) \exp\left[-1.158\left\{(1 - P^{2.35})^{-1/2.35} - 1\right\}^{0.58}\right], \tag{55}$$

that has an R equivalent,

Ibarra 2 R (IB2R):

$$s_{\mathrm{IB2R}} = \left(\pi/3^{1/2}\right) \times \exp\left[-1.158\left\{(1 - R^{2.35})^{-1/2.35} - 1\right\}^{0.58}\right]. \tag{56}$$

Considering the last four estimators, Equation 52 and Equation 54 will not be defined in the case $\overline{s}_v = 0$, while Equations 55 and 56 cannot be computed for $P = 1$ and $R = 1$ respectively, since there would be a zero in the denominator. These cases could be taken as the situation where the term in the exponential goes to minus infinity, and hence can be separately catered for by setting the estimator equal to zero. In addition, Equation 54 cannot be computed for $\overline{V}/\overline{S}_V < \frac{1}{2}\sqrt{\pi}$. This imposes a practical limit on its availability.

4. Evaluation of the assumptions used

The data used to evaluate the models were gathered by the GEWEX (Global Energy and Water Cycle Experiment) Asian Monsoon Experiment (or GAME) project between February and March 1999 and have been made available by Dr M. Toda from the website http://hydro.iis.u-tokyo.ac.jp/GAME-T/GAIN-T/field/EGAT/flux/1999Feb.html. The geographical location is in central Thailand (99°25′79″E, 16°56′38″N) at 121 m asl, where the topography is mainly flat except for a 100 m hill about 1500 m away to the east. The apparatus was set up on and around a tower 120 m high and consisting of an ultrasonic anemometer (Kaijo Corp, Model: DA-600, Probe Model: TR-90AH). The data were recorded every six seconds from a height of 30 m and the sampling time lasted 10 min.

Before proceeding to evaluate the estimators, the data were tested to see how well they conformed to the various assumptions and relations made in the derivation of the predictors. The first to be tested was the relation given by Yamartino that the persistence is approximately equal to R (Equation 24), where R was taken as the predicted and P as the predictor. Figure 1 shows the variation of R with P while Table 1 contains a summary of the statistical tests performed. The first two entries of the table refer to the correlation coefficient (r) and index of agreement (IOA) that give an indication of the amount of correlation between the predicted and predictor. Their absolute magnitude should be as close to one as possible for good agreement. All the others give a measure of the difference between the two variables. These included the mean bias (MB), the mean normalised bias (MNB), the mean absolute bias (MAB), the mean normalised absolute bias (MNAB), the root mean square error (RMSE), the normalised mean square error (NMSE), the fractional bias (FB),

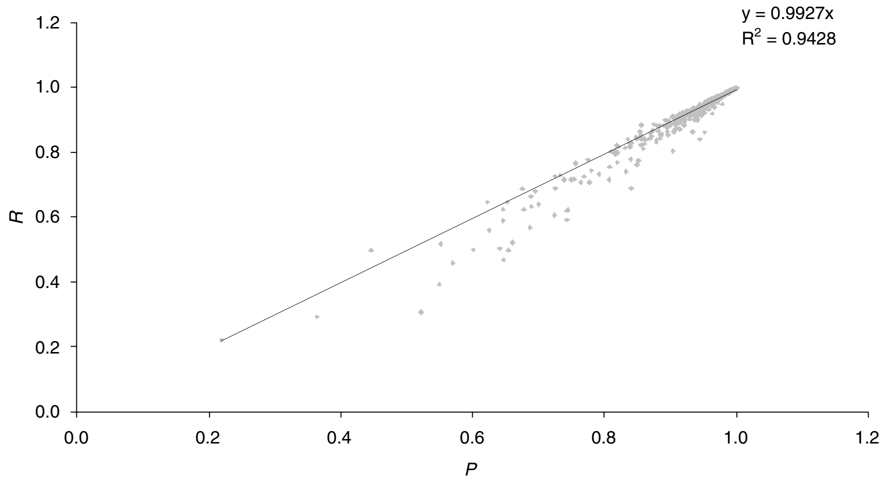

Figure 1. *Plots of R (Equation 4) against P (Equation 10).*

the maximum absolute bias (MaxAB) and the maximum point percentage normalised absolute bias (MPPNAB). Their values should be as close to zero as possible for good agreement.

Both the graph and the statistical tests indicate substantial agreement, even though there is some dispersion in the points. It is clear from Figure 1 that R and P tend to converge at values close to one, while there is a trend towards an increase in the dispersion of points at low values. The fractional bias indicates that there is a slight general over-prediction. This result is similar to that obtained by Mori (1986).

Another point that was investigated was the distribution of the wind speed, its direction and Cartesian components. For this purpose a 10-minute sample was chosen at random and analysed. Figure 2 shows the relevant histograms. The one for the wind speed might be considered to have a shape similar to the Weibull or Rayleigh distribution, the distributions that are usually used to describe it. However, the histogram for the wind direction does not seem to be unimodal, hence it should not be possible to describe it using a normal or wrapped normal distribution. As for the Cartesian components, V_x and V_y, these were factually tested for normality using the Kolmogorov-Smirnov test using a 0.5 level of significance:

$$H_0: \text{ the data is normally distributed.}$$
$$H_1: \text{ the data is not normally distributed.}$$

The results shown in Table 2 give the D statistic for V_x and V_y as 0.038 and 0.071 respectively, while the level of significance is zero in both cases. Hence H_0 is rejected and H_1 is accepted. This means that the data are not normally distributed. For V_y this is understandable since Figure 2d indicates that this variable seems to be bimodal. As for V_x the explanation might be that the distribution is not sufficiently smooth.

Table 2. *The results of the Kolmogorov-Smirnov test for normality for the Cartesian components of the wind, V_x (Equation 5) and V_y (Equation 6). Table obtained using SPSS®.*

	V_x	V_y
Number of data points	6001	6001
Normal parameters[a,b]		
Mean	−0.2666	−0.9345
Std. deviation	0.66385	0.84101
Most extreme differences		
Absolute	0.038	0.071
Positive	0.014	0.071
Negative	−0.038	−0.041
Kolmogorov-Smirnov Z	2.912	5.466
Asymp. Sig. (2-tailed)	0.000	0.000

Notes: [a]Test distribution is normal. [b]Calculated from data.

Table 1. *Results of the statistical tests performed for the relation between R (Equation 4) and P (Equation 10) as given by Equation 24.*

r	IOA	MB	MNB	MAB	MNAB	RMSE	NMSE	FB	MaxAB	MPPNAB (%)
0.9817	0.9817	0.0082	0.0124	0.0092	0.0137	0.0252	0.0007	0.0087	0.2129	69

Note: See Appendix for definition of variables and abbreviations used.

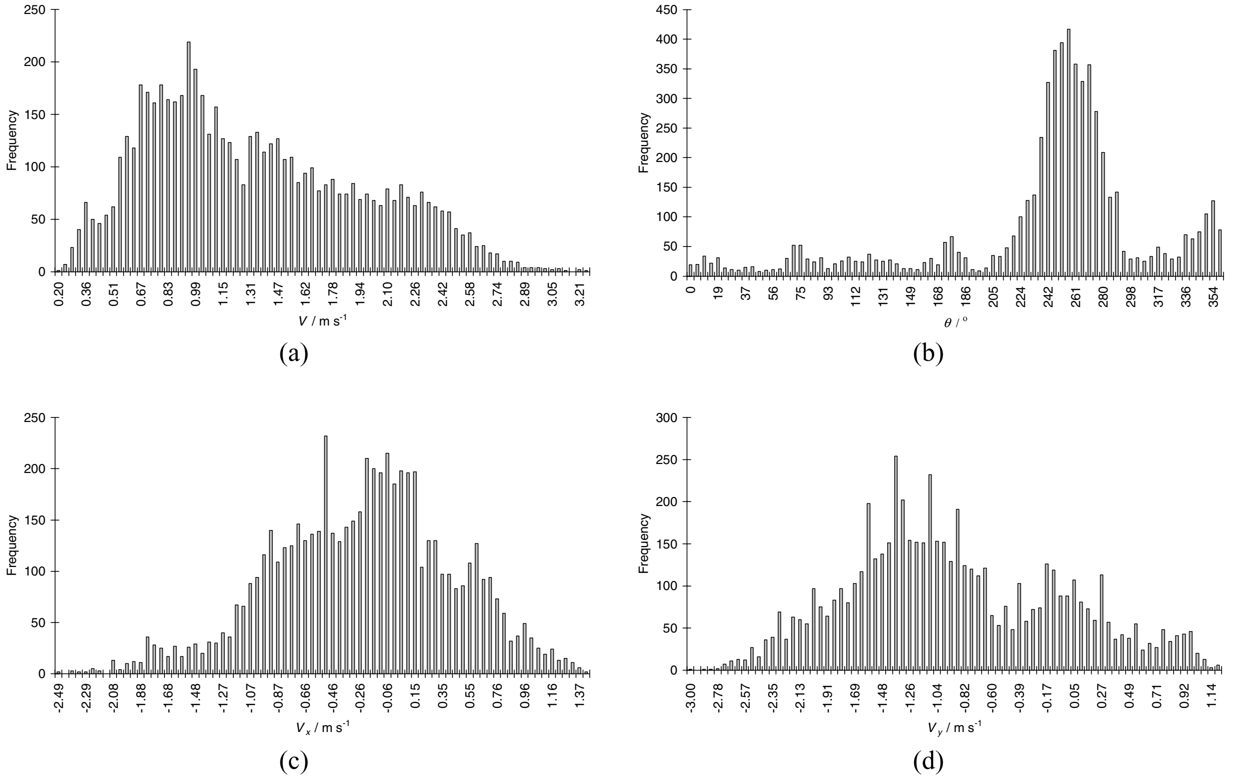

Figure 2. *The histograms of: (a) the wind speed, V, (b) the angle, θ, (c) the x-component of the wind, V_x and (d) the y-component of the wind V_y. The frequency refers to the number of points having values greater than the previous and less or equal to the value of the bin.*

5. Evaluation of the estimators

From the above, it is readily apparent that there is a relatively restricted choice of parameters on which the equations depend. Among the simplest parameters it is possible to include R, P, $\overline{V}/\overline{s}_v$, $|\overline{\mathbf{V}}|/\overline{s}_v$, and $s_{Vv}/|\overline{\mathbf{V}}|$. Their variation with s_{Y0} is shown in Figures 3–8 together with the assumed behaviour given by the various equations that depend on these parameters. For s_{VW} and s_{CVW}, which require more complex parameterisation, they were plotted directly against s_{Y0} (Figure 9).

The statistical tests are found in Tables 3–6. An additional entry was included to show the number of periods not evaluated (NPNE) due to a limit in the algorithm of the estimator.

Apart from s_{Ack}, which will be treated separately, all the estimators described show generally good agreement as

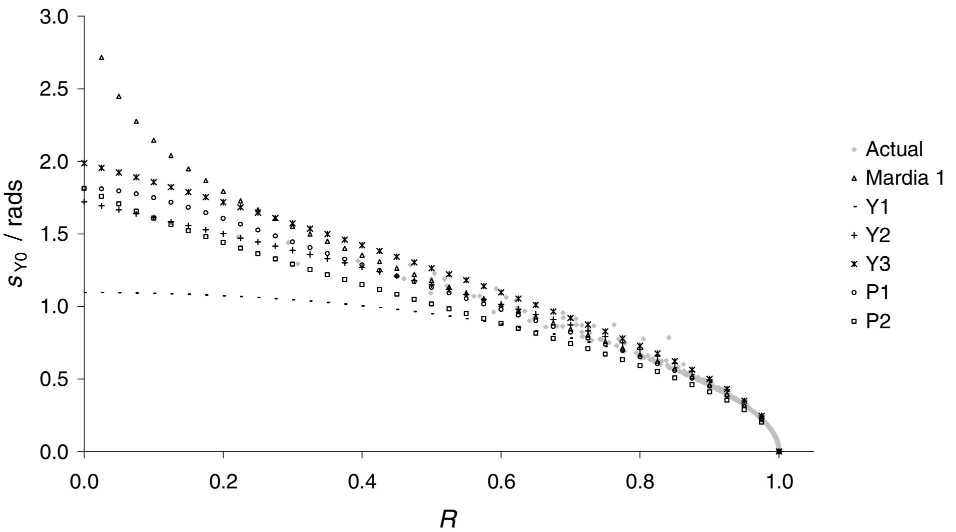

Figure 3. *The graph shows the variation of R (Equation 4) with the s_{Y0} (Equation 18). The plot also shows the assumed behaviour of R given by the estimators labelled Mardia 1 (Equation 27), Y1 (Equation 20), Y2 (Equation 21), Y3 (Equation 23), P1 (Equation 30) and P2 (Equation 31).*

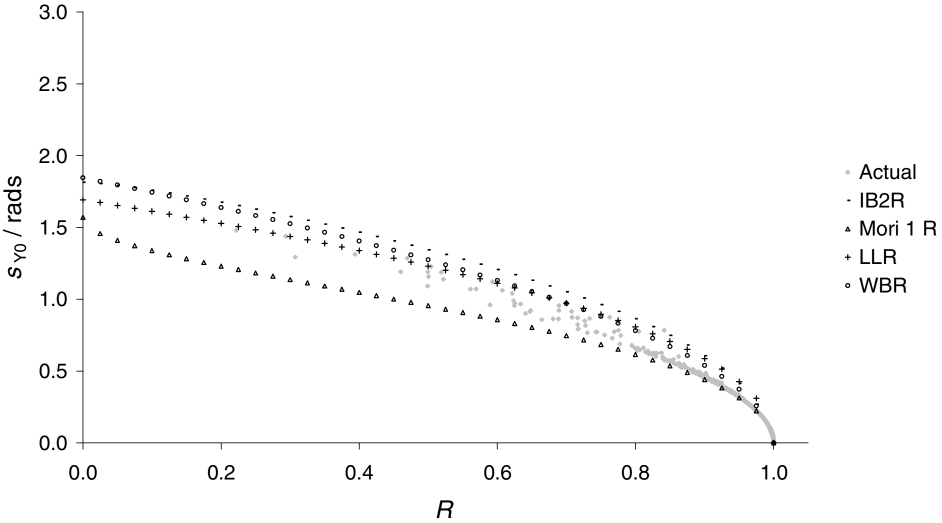

Figure 4. *The graph shows the variation of R (Equation 4) with the s_{Y0} (Equation 18). The plot also shows the assumed behaviour of R given by the estimators labelled IB2R (Equation 56), Mori 1 R (Equation 44), LLR (Equation 50) and WBR (Equation 48).*

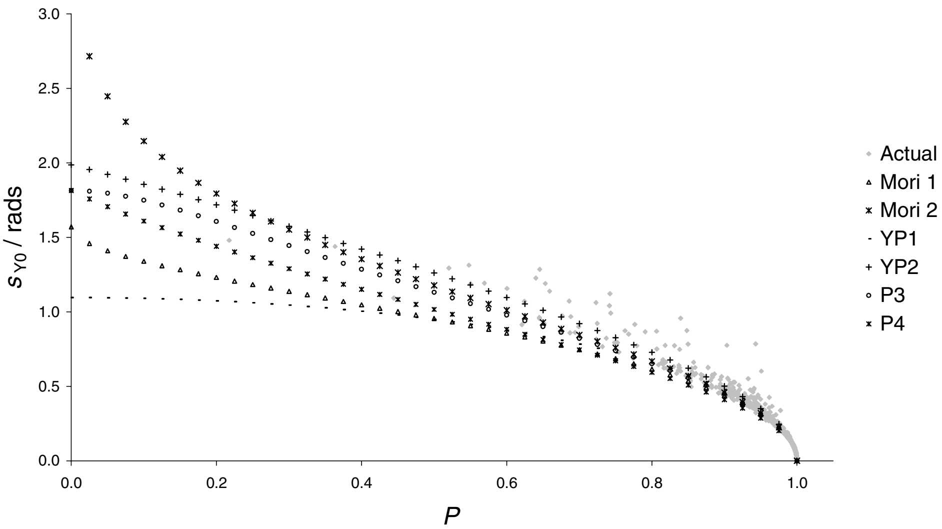

Figure 5. *The graph shows the variation of P (Equation 10) with the s_{Y0} (Equation 18). The plot also shows the assumed behaviour of P given by the estimators labelled Mori 1 (Equation 42), Mori 2 (Equation 43), YP1 (Equation 25), YP2 (Equation 26), P3 (Equation 45) and P4 (Equation 46).*

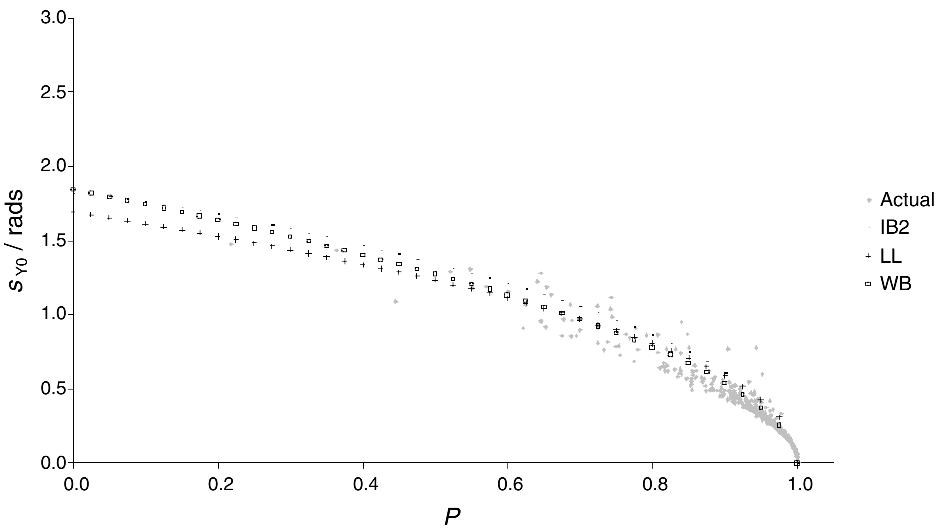

Figure 6. *The graph shows the variation of P (Equation 10) with the s_{Y0} (Equation 18). The plot also shows the assumed behaviour of P given by the estimators labelled IB2 (Equation 55), WB (Equation 47) and LL (Equation 49).*

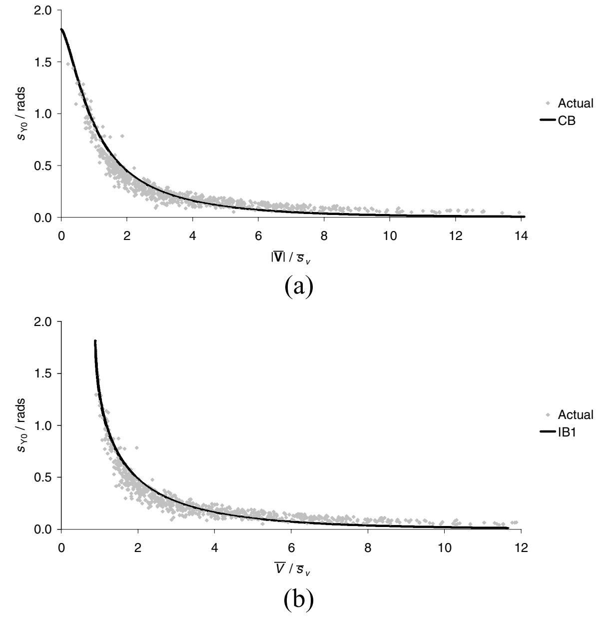

Figure 7. *(a) shows the variation of $|\mathbf{V}|/\overline{s}_v$ (Equations 8 and 53) with the s_{Y0} (Equation 18). The plot also shows the assumed behaviour of $|\mathbf{V}|/\overline{s}_v$ given by the estimators labelled CB (Equation 52); (b) shows the variation of $\overline{V}/\overline{s}_v$ (Equations 9 and 53) with s_{Y0}. The plot also shows the assumed behaviour of $\overline{V}/\overline{s}_v$ given by the estimators labelled IB1 (Equation 54).*

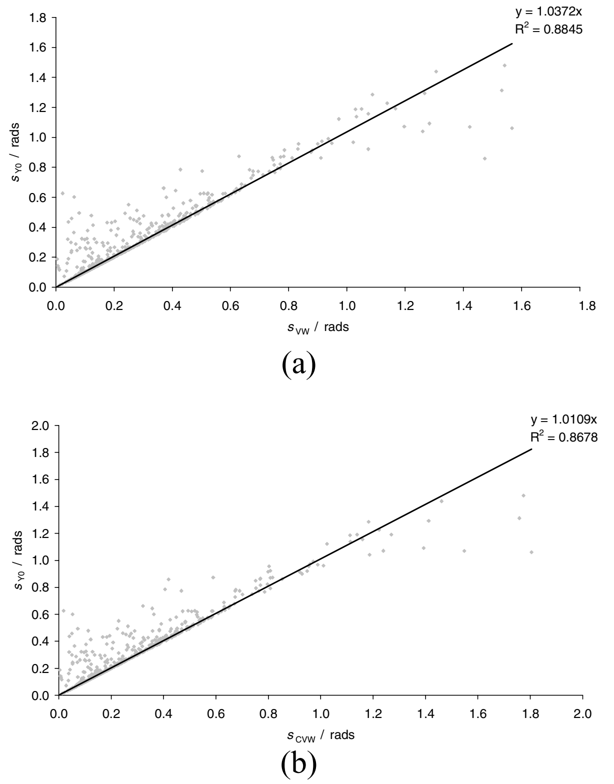

Figure 9. *(a) shows the variation of s_{Y0} (Equation 18) with s_{VW} (Equation 38); (b) shows the variation of s_{Y0} with s_{CVW} (Equation 41).*

can be seen from the values of the correlation coefficient and the index of agreement that are greater than 0.94. The measures of the errors have generally moderate values, even though these have to be related to the small magnitude of the quantities being estimated. In fact, the maximum percentage point absolute bias indicates a relative absolute over-prediction that ranged from 0.27 to 2 times the value to be estimated. The highest values were obtained for s_{CB} and s_{1B1}. Such maxima are attained at low values of dispersion where small differences can become relatively significant.

Comparing the figures for R and P, it can be seen that these parameters, and consequently the equations that depend on them, have a parallel behaviour, as would be expected from Equation 24, with R exhibiting less dispersion of points about the main trend than P. Looking at the statistical tests it immediately emerges that the measures of agreement for the equations based on R are higher, while those for dispersion favour one parameter or the other depending on the equation and/or the measure. Thus, it should be concluded that R is to be preferred. However, the difference is not particularly significant and this

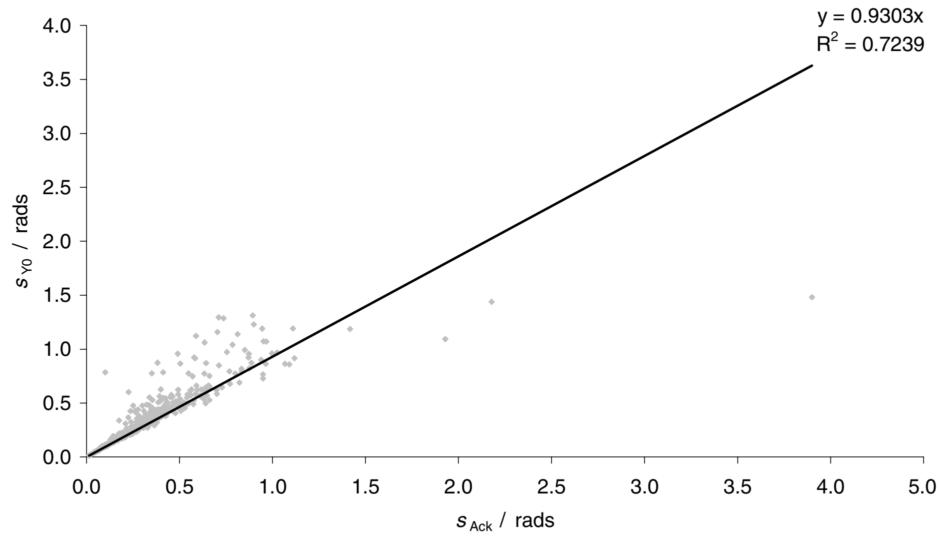

Figure 8. *The variation of s_{Y0} (Equation 18) with s_{Ack} (Equation 35).*

Table 3. *Results of the statistical tests performed for the estimators labelled Y1 (Equation 20), Y2 (Equation 21), Y3 (Equation 23), YP1 (Equation 25), YP2 (Equation 26) and Mardia 1 (Equation 27).*

	Y1	Y2	Y3	YP1	YP2	Mardia 1
NPNE	0	0	0	0	0	0
r	0.9886	0.9971	0.9977	0.9762	0.9815	0.9963
IOA	0.9783	0.9960	0.9987	0.9568	0.9842	0.9981
MB (°)	−0.0229	−0.0097	−0.0032	−0.0361	−0.0209	−0.0012
MNB	−0.0341	−0.0144	−0.0059	−0.0644	−0.0418	−0.0027
MAB (°)	0.0229	0.0097	0.0050	0.0368	0.0238	0.0052
MNAB	0.0341	0.0144	0.0080	0.0679	0.0500	0.0074
RMSE (°)	0.0642	0.0293	0.0173	0.0874	0.0573	0.0210
NMSE	0.0532	0.0106	0.0036	0.1040	0.0421	0.0053
FB	−0.0821	−0.0340	−0.0110	−0.1329	−0.0750	−0.0041
MaxAB (°)	0.5051	0.2142	0.2003	0.5218	0.4438	0.2547
MPPNAB (%)	34	27	26	58	57	25

Note: NPNE = number of periods not evaluated; see Appendix for the other abbreviations.

Table 4. *Results of the statistical tests performed for the estimators labelled P1 (Equation 30), P2 (Equation 31), Ack (Equation 35), VW (Equation 38), CVW (Equation 41) and Mori 1 (Equation 42).*

	P1	P2	Ack	VW	CVW	Mori 1
NPNE	0	0	0	0	0	0
r	0.9973	0.9974	0.8728	0.9469	0.9415	0.9800
IOA	0.9981	0.9888	0.9298	0.9698	0.9674	0.9679
MB (°)	−0.0081	−0.0331	−0.0135	−0.0278	−0.0258	−0.0321
MNB	−0.0232	−0.1075	−0.0311	−0.0819	−0.0792	−0.0589
MAB (°)	0.0091	0.0331	0.0404	0.0349	0.0342	0.0329
MNAB	0.0241	0.1075	0.0792	0.0889	0.0870	0.0629
RMSE (°)	0.0210	0.0484	0.1307	0.0838	0.0886	0.0772
NMSE	0.0054	0.0314	0.2133	0.0925	0.1026	0.0797
FB	−0.0283	−0.1212	−0.0479	−0.1009	−0.0931	−0.1174
MaxAB (°)	0.2106	0.2618	2.4200	0.6151	0.7435	0.4768
MPPNAB (%)	27	33	163	98	98	57

Note: NPNE = number of periods not evaluated; see Appendix for the other abbreviations.

Table 5. *Results of the statistical tests performed for the estimators labelled Mori 2 (Equation 43), Mori 1 R (Equation 44), P3 (Equation 45), P4 (Equation 46), WB (Equation 47) and WBR (Equation 48).*

	Mori 2	Mori 1 R	P3	P4	WB	WBR
NPNE	0	0	0	0	0	0
r	0.9794	0.9949	0.9808	0.9810	0.9820	0.9972
IOA	0.9845	0.9892	0.9816	0.9627	0.9892	0.9893
MB (°)	−0.0195	−0.0174	−0.0256	−0.0489	0.0199	0.0398
MNB	−0.0392	−0.0270	−0.0586	−0.1395	0.0838	0.1255
MAB (°)	0.0234	0.0174	0.0272	0.0489	0.0324	0.0401
MNAB	0.0489	0.0270	0.0618	0.1395	0.1019	0.1262
RMSE (°)	0.0570	0.0467	0.0613	0.0839	0.0508	0.0536
NMSE	0.0415	0.0276	0.0490	0.1008	0.0287	0.0300
FB	−0.0696	−0.0620	−0.0924	−0.1841	0.0662	0.1285
MaxAB (°)	0.4424	0.3020	0.4493	0.4782	0.3849	0.2234
MPPNAB (%)	56	30	57	61	49	21

Note: NPNE = number of periods not evaluated; see Appendix for the other abbreviations.

means that it should be viable to obtain and use P from databases that contain only R and vice versa. This was in fact the reason why Yamartino proposed Equation 24. With such considerations in mind, it is possible to restrict the comparative analysis of the performance of the estimators to those depending on R, since those depending on P will follow automatically.

Table 6. *Results of the statistical tests performed for the estimators labelled LL (Equation 49), LLR (Equation 50), CB (Equation 52), IB1 (Equation 54), IB2 (Equation 55) and IB2R (Equation 56).*

	LL	LLR	CB	IB1	IB2	IB2R
NPNE	0	0	0	0	0	0
r	0.9790	0.9937	0.9706	0.9642	0.9785	0.9926
IOA	0.9733	0.9709	0.9785	0.9767	0.9798	0.9727
MB (^{c})	0.0620	0.0805	0.0209	0.0183	0.0319	0.0537
MNB	0.3305	0.3715	−0.0367	−0.0391	0.0040	0.0520
MAB (^{c})	0.0711	0.0807	0.0602	0.0601	0.0532	0.0677
MNAB	0.3406	0.3717	0.2928	0.2968	0.2117	0.2381
RMSE (^{c})	0.0794	0.0876	0.0765	0.0786	0.0747	0.0918
NMSE	0.0618	0.0714	0.0650	0.0692	0.0598	0.0846
FB	0.1931	0.2438	0.0695	0.0611	0.1043	0.1695
MaxAB (^{c})	0.3318	0.1661	0.2975	0.3432	0.3390	0.2750
MPPNAB (%)	65	71	200	210	99	98

Note: NPNE = number of periods not evaluated; see Appendix for the other abbreviations.

It can be noted from Figure 3 that, as stated by Yamartino (1984), s_{Y1} underestimates the real trend, while s_{Y2} and s_{Y3} represent successive improvements to the original equation. Over-prediction is also exhibited by $s_{Mardia\ 1}$, which as expected, diverges to infinity for small values of R. However both s_{P1} and s_{P2} adjust for this behaviour of $s_{Mardia\ 1}$. Unfortunately, the limited number of data point in the region of high angular dispersion does not allow a proper analysis of the behaviour of $s_{Mardia\ 1}$.

Regarding the other equations parameterised by R, Figure 4 indicates that $s_{Mori\ 1R}$ underestimates, while s_{IB2R}, s_{WBR} and s_{LLR} slightly overestimate the real trend. Comparing s_{WB} and s_{LL}, the statistical tests performed (Tables 5 and 6) were slightly in favour of the former, even though the differences were small. It is plausible that the discrepancy in the constants between the two equations resulted from the differences in the data sets used by Webber (1992) and Leung & Liu (1996). This is to be expected with data fitting since the various factors that affect the wind mean that samples taken at different times and locations are statistically very different from one another.

The results obtained using the isotropic Gaussian model give a reasonable outcome. Both $|\overline{\mathbf{V}}|/\overline{s}_{VIGM}$ (Figure 7a) and $\overline{V}/\overline{s}_{V\ IGM}$ (Figure 7b) show a good trend with s_{Y0}, something that is coupled with small scattering in the data points. The existing equations that employ these as parameters, which are respectively given by s_{CB} and s_{IB1}, manage to reproduce the actual trend relatively well. The major problem is shown by the high values of the maximum point percentage absolute bias that is very high. This is probably due to discrepancies between the estimated and the actual value encountered in regions of small angular dispersion where small differences become relatively significant.

Moving to the estimator developed by Verrall & Williams (1982) (Equation 38) and its corrected form (Equation 41), Figure 9 shows that the range of scattering of the actual points is only contained in regimes of medium angular dispersion. This, however, did not result in an excessive amount of relative absolute over-prediction as could be seen from the values of the maximum percentage point normalised absolute bias (Table 4). The statistical results also indicate that there is no significant progress made in going from s_{VW} to s_{CVW} since improvement in some quantities is offset by a deterioration in others.

As regards the performance of equations based on $s_{Vv}/|\overline{\mathbf{V}}|$, from Figure 8 it is evident that s_{Ack} describes the performance of this parameter only in the low dispersion region. The problem arises due to the fact that this estimator consists of a ratio of two statistics that, from the results obtained, do not vary linearly with one another. Thus the denominator may converge to zero faster than the numerator, leading to substantial over-prediction. Similar considerations were put forward by Yamartino (1984) even though in his evaluation this predictor did not perform so badly. Thus this predictor was not considered as a suitable estimator of s_{Y0}.

6. Conclusion

This work has examined the applicability of equations derived from theoretical considerations as estimators for the standard deviation of wind direction. It was clear that for these data the distribution of the angle for a sample was not necessarily unimodal. Thus estimators developed from the normal and wrapped normal distribution should not have been used. However, this did not give major problems in the performance of $s_{Mardia\ 1}$, s_{VW} and s_{CVW}. In addition, the Cartesian components of the wind were not found to be normally distributed, which means that the isotropic Gaussian model should not be adopted. Even so, all the estimators derived from this model, s_{CB}, s_{IB1}, s_{IB2} and s_{IB2R}, did not seem to suffer any major limitations.

During the analysis it has also been shown that most of the existing estimators considered here perform on approximately the same level. In the main there are few criteria that can help discriminate. The high relative absolute bias obtained by s_{IB1}, s_{CB} and s_{Ack} immediately advises against their use. In particular, for s_{IB1}, it might also not be possible to compute the algorithm for every data set, while for s_{Ack}, there seems to be a trend towards increasing divergence for increasing angular dispersion. Regarding s_{CVW} and s_{VW}, these graphically show extensive dispersion of points, which is backed up by a substantially high relative error and low measures of agreement. This again suggests that they should not be adopted.

The parameter that shows the best trend together with the least dispersion of points is R. In fact the best statistical results were obtained by equations based on R. These include, in order of decreasing performance, s_{Y3}, s_{WBR}, s_{Mardia} and s_{P1}. Such an outcome might have been expected since the approximation of s_{Y0} made by Yamartino (1984) immediately relates his algorithm to R.

Here it ought to be mentioned that due to a different concept in the estimation of the standard deviation, $s_{Mardia\ 1}$ might diverge at high values of angular dispersion. It has been shown that such a limit can be overcome by using s_{P1} and s_{P2}. Thus, if the standard deviation of wind direction is considered to be bounded, estimators that diverge should be put aside in favour of those that have an upper limit.

The analysis performed here for the R estimators was also valid for their P equivalent. It was shown that Equation 24 holds to a reasonable degree. The explanation for any discrepancy can be found in the fact that P is meant to account for 'vectorial' variation of the wind direction while R takes into account only the value of the angle. Since wind is a vector quantity it should follow that the persistence represents much better the angular dispersion than R. The reason why R has a better trend than P is simply due to the fact that s_{Y0} considers only the angular and not the 'vectorial' variation of the wind direction. Such treatment of wind statistics can be extended beyond the angular standard deviation to the mean angle and wind speed. There is some concern whether the statistical measures that are being adopted are adequate enough to describe the wind samples. Such a question will require further investigation.

Acknowledgements

We would like to thank Dr M. Toda and GAME for providing the data that have been used to evaluate the equations.

References

Ackermann, G. R. (1983) Means and standard deviation of horizontal wind components. *J. Climate & Appl. Meteorol.* **22**: 959–961.

Baird, D. C. (1962) *Experimentation, An Introduction to Measurement Theory and Experiment Design*, Prentice-Hall.

Castans, M. & Barquero, C. G. (1994) A framework for the structure of a low wind speed field. *Bound.-Layer Meteor.* **69**: 137–147.

Castans, M. & Barquero, C. G. (1998) Some comment on the study of low persistence wind: discussion. *Atmos. Environ.* **32**: 253–256.

Essenwanger, O. M. (1985) *World Survey of Climatology Volume 1 B: General Climatology*, 1 B: *Elements of Statistical Analysis*. Amsterdam: Elsevier.

Fisher, N. I. (1983) Comment on 'A method for estimating the standard deviation of wind direction'. *J. Climate & Appl. Meteorol.* **22**: 1971–1972.

Ibarra, J. I. (1995) A new approach for the determination of horizontal wind direction fluctuations. *J. Climate & Appl. Meteorol.* **34**: 1942–1949.

Leung, D. Y. C. & Liu, C. H. (1995) Improved estimators for the standard deviations of horizontal wind fluctuations. *Atmos. Environ.* **30**: 2457–2461.

Mardia, K. V. (1972) *Statistics of Directional Data*. New York: Academic Press.

Mori, Y. (1986) Evaluation of several 'single-pass' estimators of the mean and the standard deviation of wind direction. *J. Climate & Appl. Meteorol.* **25**: 1387–1386.

Verrall, K. A. & Williams, R. L. (1982) A method for estimating the standard deviation of wind direction. *J. Climate & Appl. Meteorol.* **21**: 1922–1925.

Weber, R. (1992) A comparison of different estimators for the standard deviation of wind direction based on persistence. *Atmos. Environ.* **26**: 983–986.

Yamartino, R. J. (1984) A comparison of several 'single pass' estimators of the standard deviation of wind direction. *J. Climate & Appl. Meteorol.* **23**: 1362–1366.

Appendix: Symbols and abbreviations used in the text and tables

$\overline{C}$: the mean of the cosines of the angles (Equation 1)
$\Delta(\phi)_i$: gives the minimal angular distance between θ_i and ϕ (Equation 19)
FB: the fractional bias
IOA: the index of agreement
MaxAB: the maximum absolute bias
MAB: the mean absolute bias
MB: the mean bias
MNAB: the mean normalised absolute bias
MNB: the mean normalised bias
MPPNAB: maximum point percentage normalised absolute bias
n: number of angles in a data set
NMSE: the normalised mean square error
μ_X: the population mean of the variable X
P: the persistence (Equation 10)
θ: a general angle
θ_a: the angle the mean wind vector to the centre of gravity of the system, **R**, makes with the positive

x-axis (Equation 3). It is taken to be a measure of the mean angle

θ_i: one of the angles in a data set, $i = 1, 2, \ldots n$

θ_v: the angle the mean wind vector, $\overline{\mathbf{V}}$, makes with the positive x-axis (Equation 7). It is taken to be the mean vector angle

r: the correlation coefficient

R: the length of the vector to the centre of gravity of the system (Equation 4)

R: the vector to the centre of gravity

RMSE: the root mean square error

$\overline{S}$: the mean of the cosines of the angles (Equation 2)

s_{Ack} and Ack: the estimator given by Ackermann (1983) (Equations 34 and 35)

s_{C}^2 and s_{C}: respectively the sample variance and standard deviation of the cosine of the angles (Equation 11)

s_{CB}, Castans & Barquero and CB: the estimator presented by Castans & Barquero (1994) (Equation 52)

s_{CVW}, Corrected Verrall & Williams and CVW: the corrected version of the estimator presented by Verrall & Williams (1982) proposed by Yamartino (1984) (Equation 41)

s_{LL}, Leung & Liu and LL: the estimator given by Leung & Liu (1995) (Equation 49)

s_{LLR}, Leung & Liu R and LLR: the R equivalent of the estimator given by Leung & Liu (1995) (Equation 50)

$s_{\mathrm{IB}i}$, Ibarra i, and IBi: the ith estimator presented by Ibarra (1995) (Equations 54 and 55)

s_{IB2R}, Ibarra 2 R, and IB2R: the R equivalent of the second estimator presented by Ibarra (1995) (Equation 56)

$s_{\mathrm{Mardia\ 1}}$ and Mardia 1: the estimator given by Mardia (1972) (Equation 27)

$s_{\mathrm{Mori}\,i}$ and Mori i: the ith presented by Mori (1986) (Equations 42 and 43)

$s_{\mathrm{Mori\ 1\ R}}$ and Mori 1 R: The R version of the first estimator presented by Mori (1986) (Equation 44)

$s_{\mathrm{P}i}$ and Predictor i: the ith estimator presented in this work (Equations 30, 31, 45, 46)

s_{S}^2 and s_{S}: respectively the sample variance and standard deviation of the sine of the angles (Equation 12)

s_{Vu}^2 and s_{Vu}: respectively the sample variance and standard deviation of the components of the vector wind speed along the u-axis (Equation 16)

s_{Vv}^2 and s_{Vv}: respectively the sample variance and standard deviation of the components of the vector wind speed along the v-axis (Equation 17)

s_{VW}, Verrall & Williams and VW: the estimator given by Verrall & Williams (1982) (Equation 38)

s_{Vx}^2 and s_{Vx}: respectively the sample variance and standard deviation of the components of the vector wind speed along the x-axis (Equation 13)

s_{Vy}^2 and s_{Vy}: respectively the sample variance and standard deviation of the components of the vector wind speed along the y-axis (Equation 14)

s_{Vxy}^2 the x-y-velocity covariance of the vector wind speed components (Equation 15)

$\bar{s}_V$: the sample estimator of $\overline{\sigma}_V^2$ (Equation 53)

s_{WB}, Weber and WB: the estimator given by Weber (1992) (Equation 47)

s_{WBR}, Weber R and WBR: the R equivalent of the estimator given by Weber (1992) (Equation 48)

s_{Y0}: the equation given by Yamartino (1984) as the measure of the standard deviation of wind direction (Equation 18)

$s_{\mathrm{Y}i}$, Yamartino i and Yi: the ith ($i \neq 0$) estimator proposed by Yamartino (1984) (Equation 20, 21 and 23)

s_{YP1}, Yamartino P 1 and YP 1: the first estimator proposed by Yamartino (1984) given in terms of the persistence (Equation 25)

s_{YP2}, Yamartino P 2 and YP 2: the third estimator proposed by Yamartino (1984) given in terms of the persistence (Equation 26)

$\overline{\sigma}_V^2$: by definition is equal to $2\sigma_{Vx}^2 = 2\sigma_{Vy}^2$

σ_X^2 and σ_X: respectively the population variance and standard deviation of the variable X

u-axis: the x-axis rotated by an angle of θ_{V}

v: a particular value of the wind speed for the population

v-axis: the y-axis rotated by an angle of θ_{V}

$\mathbf{v}_1$ and $\mathbf{v}_2$: two general vectors

$\overline{\mathbf{V}}$: the mean wind vector of the sample (*cf*. Equations 5 and 6)

$|\overline{\mathbf{V}}|$: the magnitude of the mean wind vector of the sample (Equation 8)

V_i: the ith value of the wind speed for the sample

v_x and v_y: the component the mean wind vector of the population makes with the x- and y-axis respectively

$\overline{V}$: the mean wind speed of the sample (Equation 9)

$\overline{V}_x$ and $\overline{V}_y$: the component the mean wind vector of the sample makes respectively with the x- and y-axis (Equations 5 and 6)