Technical University of Denmark

DTU

Field calibration of cup anemometers

Kristensen, Leif; Jensen, G.; Hansen, A.; Kirkegaard, P.

Publication date:
2001

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

DTU Library
Technical Information Center of Denmark

Field Calibration of Cup Anemometers

Leif Kristensen, Gunnar Jensen,
Arent Hansen, and Peter Kirkegaard

Abstract   An outdoor calibration facility for cup anemometers, where the signals from 10 anemometers of which at least one is a reference can be can be recorded simultaneously, has been established. The results are discussed with special emphasis on the statistical significance of the calibration expressions. It is concluded that the method has the advantage that many anemometers can be calibrated accurately with a minimum of work and cost. The obvious disadvantage is that the calibration of a set of anemometers may take more than one month in order to have wind speeds covering a sufficiently large magnitude range in a wind direction sector where we can be sure that the instruments are exposed to identical, simultaneous wind flows. Another main conclusion is that statistical uncertainty must be carefully evaluated since the individual 10 minute wind-speed averages are not statistically independent.

Contents

1 Introduction

The most accurate instrument for measuring the mean-wind speed has so-far been the cup anemometer. It was invented in 1846 by the Irish astronomer T.R. Robinson (Middleton 1969, Wyngaard 1981) and the original instruments had four cups. Until the end of the 1920s much research went into experimenting with the number of cups and the arm lengths. Brazier (1914) and Patterson (1926) found that shorter arms improved the linearity of the calibration and that a three-arm cup rotor is optimal with respect to sensitivity and suppression of the unevenness in the rotation ("wobbling"). A modern cup anemometer is shown in Fig. 1.
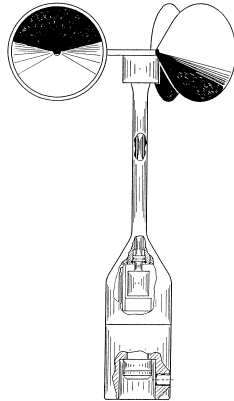

Figure 1. The Risø anemometer. The diameter of the conical cups in the three-cup plast rotor is 7 cm. The body is made of anodized aluminum. The height of the instrument from the bottom to the center of the rotor is 24.7 cm. The output signal, generated by a magnetically activated switch, is a train of electric pulses, two for each rotor revolution. By counting pulses over a certain period one obtains a number corresponding to the mean-wind speed for this period.

The cup anemometer now appears to be the preferred instrument for measuring the mean-wind speed in the wind-energy community, mainly because of the linearity, the accuracy and the stability of the calibration. But also the fact that this instrument is omnidirectional and easy to mount makes it attractive for routine measurements. There has been some discussion about the importance of the so-called "overspeeding" which is caused by the asymmetric response to instantaneous increases and decreases in the wind speed and which, incidentally, is a necessary property for the cup anemometer to start rotating when exposed to a wind, i.e. to function at all (Kristensen 1998). Some of those who believe that overspeeding is a problem have preferred to use a propeller-vane instrument which is considered symmetric in its response to opposite wind directions. The problem is here that the propeller signal depends on the instantaneous angle between the wind direction and the propeller axis. The vane of the instrument, which tries to align the axis with the wind direction, will always lag behind. The result is that there will always be a systematic error of the measured mean wind. This error is in general exacerbated by the fact that the propeller will perform its own motion since it is seldom mounted directly over the vertical axis of the vane (Kristensen 1994). Also, on basis of the studies by Kristensen (1998), Kristensen (1999), and Kristensen (2000) the adverse effects of the asymmetric response of the cup anemometer seem exaggerated. Thus the choice to use cup anemometers for routine measurements is a sound one.

In the following we discuss an outdoor calibration setup for a number of cup anemometers which are simultaneously exposed to the same wind. In principle we can let one of

the anemometers be "the standard" and then intercalibrate by comparing the output signals. This standard anemometer is assumed to have been calibrated in a "certified" way according to an approved, well-described method by means of a good wind tunnel with little turbulence and a flat profile. It could be argued that an anemometer mounted outdoor is exposed to wind flows which changes instantaneously in both wind speed and wind direction whereas the flow in a wind tunnel is always in the same direction. However, since the cup anemometer allegedly measures the instantaneous component[*] we really compare mean signals from these outdoor anemometers through which the same "length of air" has passed, just as we would have done when using a wind tunnel. The only reservation one can have is that the non-ideal angular response to vertical wind components may produce a bias due to the fluctuating vertical wind component, but, as demonstrated by Kristensen (1994), this can relatively easily be quantified if necessary. Here it is not considered a serious problem.

First we describe the calibration site. Then we discuss the statistical data analysis and, finally, we illustrate the method by analyzing real data from the calibration installation.

2 Calibration Setup

The calibration setup is located behind the beach at the southwestern part of the Risø peninsula, just north of the pier, as Fig. 2 shows. The distance to the water line is approximately 12 m and the orientation of the calibration boom where the ten anemometers are mounted about 10 m over the surrounding terrain is 19°–199°. This means that wind from the direction 289° will have travelled over a water fetch of about 7 km before it simultaneously reaches all the anemometer. Accepting a direction sector of ±45° around 289°, the water fetch will be at least 3 km. In other words, the site has been chosen to be well-exposed for winds from the predominant wind direction in Denmark.

The calibration boom is mounted on the top of two steel lattice masts with triangular cross sections with the side length 0.25 m. Fig. 3 shows a sketch and a photograph of the measuring setup.

The data recording is carried out with an Aanderaa datalogger with 12 ten-bit channels. Numbering these from 1 to 12, channel 1 is the temperature channel and channel 12 the direction channel. Channels from 2 to 11 are used for the anemometer outputs. To prevent the signal from the Risø anemometer with two counts per revolution from causing the 10-bit channels to overflow during the counting period of 10 minutes, the numbers in the registers are scaled down by the factor $2^5 = 32$. Usually we have two positions for reference anemometers. The rest of the positions are test positions. Table 1 gives the "names" of the ten positions.

This calibration configuration, where uninterrupted, consecutive 10-minute averages have been measured, has been operating since 1996.

3 Statistical Considerations

The calibration procedure implies that we compare two almost identical responses to the same signal. For symmetry reasons we adopt a method where the mean square of the

---

[*]Actually, averaged over one full rotor rotation.
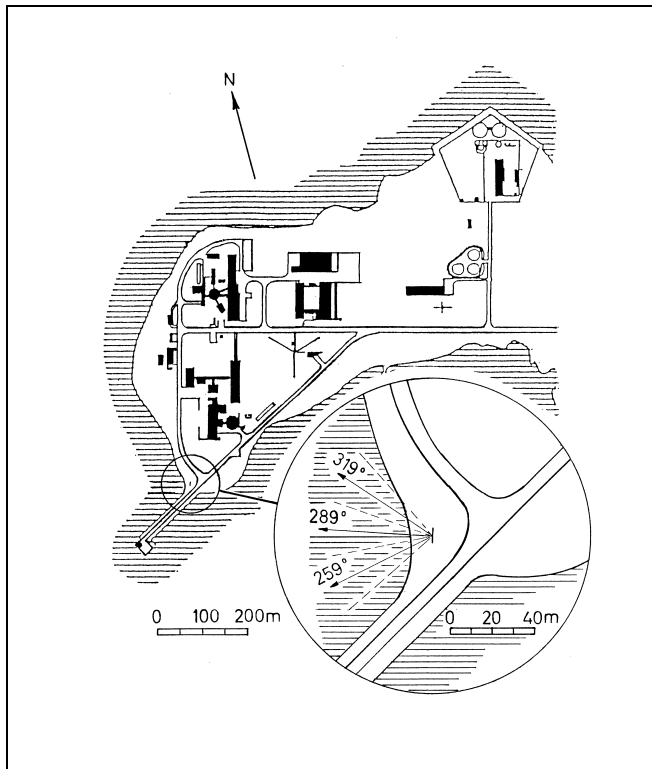
Figure 2. Map of the Risø peninsula. The calibration site is shown in a little more detail in the circular blow-up. The direction perpendicular to the calibration boom, which parallel to the coastline, is 289°. When the wind is from this direction the mutual flow disturbance between the anemometers is at minimum.
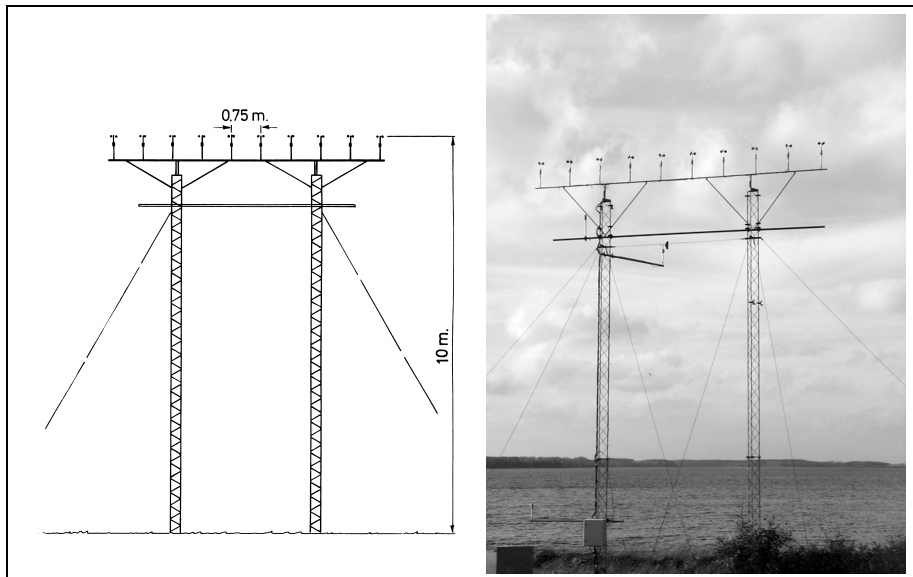
Figure 3. Calibration setup at Risø, in the left frame drawn to the scale 1:190. The orientation of the boom with the ten cup anemometers is 19°–199°. The lower boom is a service boom used for mounting the instruments. The wind direction is measured at the end of a 1.8 m long boom, pointing in the direction 319°. The thermometer is mounted on a boom at 2 m over the ground. The two supporting masts are triangular lattice masts with the side length 25 cm.

Table 1. Position names and channel numbers. The northernmost and southernmost positions are " Test A1" and " Test A2", respectively.

Position	Channel
Test A1	11
Test B1	10
Ref. 1	9
Test C1	8
Test D1	7
Test D2	6
Test C2	5
Ref. 2	4
Test B2	3
Test A2	2

perpendicular distance of the points to the regression line is minimized. Subsequently we discuss the statistical implications of the adopted method.

3.1   Gain and Offset by Orthogonal Fitting

We want to fit a straight line ℓ to a number of points (x_i, y_i), $i = 1 \ldots N$.

The equation for the line is

$$y = ax + b \tag{1}$$

or, in vector form,

$$r = r_0 + \theta t, \quad -\infty < \theta < \infty \tag{2}$$

where r_0 is a point on the line and t a unit vector in the direction of the line.

The two signals (x_i, y_i) are both representing wind velocities from the same type of anemometer. Now we minimize the mean squared sum ϕ of the *perpendicular* distances d_i from the points to the line, viz.

$$\phi = \frac{1}{N} \sum_{i=1}^{N} d_i^2. \tag{3}$$

We must therefore first determine the distance d_i from a point (x_i, y_i) to the line ℓ.

In terms of the unit vectors i and j describing the Cartesian coordinate system we have

$$r_0 = x_0 i + y_0 j. \tag{4}$$

The unit vector t along ℓ and the orthogonal unit vector n are given by

$$t = \quad \cos \alpha \, i + \sin \alpha \, j \tag{5}$$

and

$$n = -\sin\alpha\, i + \cos\alpha\, j, \tag{6}$$

where α is the angle from the x-axis to ℓ.

The signed distance d_i from a given point

$$r_i = x_i i + y_i j \tag{7}$$

to the line ℓ is

$$d_i = (r_i - r_0) \cdot n = -(x_i - x_0)\sin\alpha + (y_i - y_0)\cos\alpha. \tag{8}$$

The definitions of the quantities we use are illustrated in Fig. 4.
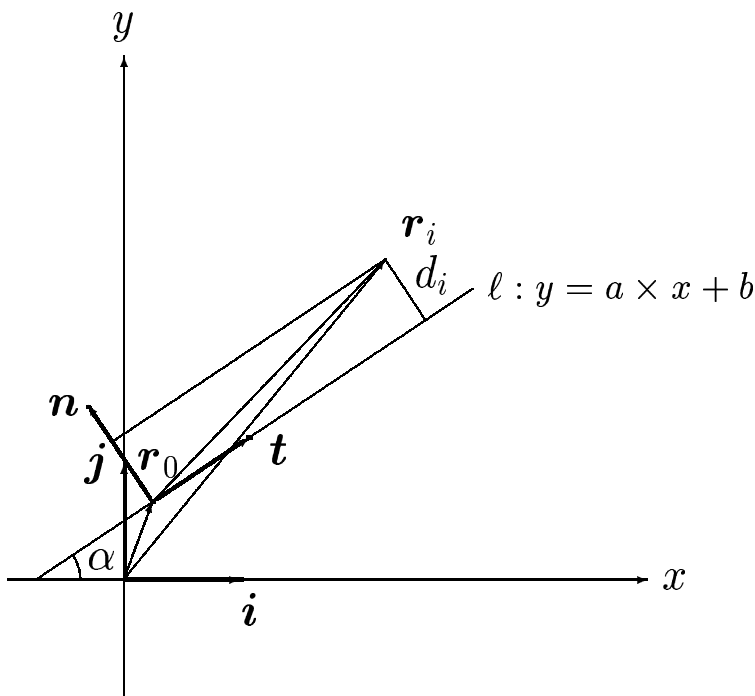
Figure 4. Illustration of the calculation of the distance d_i of one point to the regression line with offset b and slope $a = \tan\alpha$.

Equation (3) now becomes

$$\phi = \frac{1}{N}\sum_{i=1}^{N}\left\{ -(x_i - x_0)\sin\alpha + (y_i - y_0)\cos\alpha \right\}^2. \tag{9}$$

We want to determine the values of r_0 and α that minimize ϕ. In fact, there are only two independent parameters of which the angle α must be one. Since we expect α to be different from $n \times \pi/2$, we can fix either of x_0 or y_0 and use the other as a fitting parameter. Let us fix x_0. In principle, there is no bounds on its value, but it seems practical to choose $x_0 = \bar{x}$ where the average symbol stands for the operation

$$\bar{z} = \frac{1}{N}\sum_{i=1}^{N} z_i. \tag{10}$$

To find y_0 we demand

$$
\begin{aligned}
0 = \frac{\partial \phi}{\partial y_0} &= \frac{1}{N} \sum_{i=1}^{N} 2\{-(x_i - x_0)\sin\alpha + (y_i - y_0)\cos\alpha\}(-\cos\alpha) \\
&= 2\cos\alpha\{+(\bar{x} - x_0)\sin\alpha - (\bar{y} - y_0)\cos\alpha\} \\
&= -2(\bar{y} - y_0)\cos^2\alpha
\end{aligned}
\tag{11}
$$

which implies that $y_0 = \bar{y}$.

For practical reasons we introduce new coordinates by moving the origin to $(\bar{x}, \bar{y})$ and just make the following replacements:

$$
\left\{ \begin{array}{c} x_i \\ y_i \end{array} \right\} := \left\{ \begin{array}{c} x_i - \bar{x} \\ y_i - \bar{y} \end{array} \right\},
\tag{12}
$$

so that

$$
\phi = \frac{1}{N} \sum_{i=1}^{N} \left\{ -x_i \sin\alpha + y_i \cos\alpha \right\}^2 .
\tag{13}
$$

To determine α we demand that $\partial\phi/\partial\alpha = 0$:

$$
\begin{aligned}
0 &= \frac{2}{N} \sum_{i=1}^{N} \left\{ -x_i \sin\alpha + y_i \cos\alpha \right\} \left\{ -x_i \cos\alpha - y_i \sin\alpha \right\} \\
&= \frac{2}{N} \sum_{i=1}^{N} \left\{ (x_i^2 - y_i^2) \cos\alpha \sin\alpha - x_i y_i (\cos^2\alpha - \sin^2\alpha) \right\} \\
&= 2 \left\{ (\overline{x^2} - \overline{y^2}) \cos\alpha \sin\alpha - \overline{xy} (\cos^2\alpha - \sin^2\alpha) \right\}.
\end{aligned}
\tag{14}
$$

The solution to (14) is

$$
\tan 2\alpha = \frac{2\overline{xy}}{\overline{x^2} - \overline{y^2}}
\tag{15}
$$

or

$$
\tan\alpha = \frac{1}{2\overline{xy}} \left\{ \overline{y^2} - \overline{x^2} \pm \sqrt{\left(\overline{y^2} - \overline{x^2}\right)^2 + 4\overline{xy}^2} \right\}.
\tag{16}
$$

We see that there are two solutions and that their product is -1. This corresponds to slopes which minimize ($+$ solution) and maximize ($-$ solution) ϕ and which pertain to lines which are perpendicular to one another. We must use the $+$ solution.

Introducing

$$
\delta = \frac{1}{2} \ln \left(\frac{\overline{x^2}}{\overline{y^2}} \right)
\tag{17}
$$

and

$$\varepsilon = \frac{1}{2}\ln\left(\frac{\overline{x^2}\,\overline{y^2}}{\overline{xy}^2}\right) = -\ln(|\rho|),$$ (18)

where

$$\rho = \frac{\overline{xy}}{\left(\overline{x^2}\,\overline{y^2}\right)^{1/2}}$$ (19)

is the sample correlation coefficient, then the solution (16) can be written

$$a \equiv \tan\alpha = \text{sign}(\overline{xy})\left(\sqrt{1 + (\mathrm{e}^{\varepsilon}\sinh\delta)^2} - \mathrm{e}^{\varepsilon}\sinh\delta\right).$$ (20)

and

$$b = y_0 - a x_0 = \overline{y} - a\overline{x}.$$ (21)

In appendix A we generalize the approach to allow for a slight curvature.

3.2   Statistical Uncertainties

In the following we will make repeated use of the so-called error-propagation formula and, to set the stage and define the notation, we will reiterate its meaning and content. Then we will determine the statistical uncertainties of a and b and functions of these two quantities. Finally, we will determine if all the N realizations can be considered statistically independent in the present context or whether we must reduce the degrees of freedom.

Error Propagation

Let

$$z = (z_1, z_2, \ldots, z_M)$$ (22)

be one realization of a set of M ordered, random variables which may or may not be inter-correlated.[†]

The ensemble averages and variances of the M variables of z are $(\langle z_1 \rangle, \langle z_2 \rangle, \ldots, \langle z_M \rangle)$ and $\left(\langle (z_1 - \langle z_1 \rangle)^2 \rangle, \langle (z_2 - \langle z_2 \rangle)^2 \rangle, \ldots, \langle (z_M - \langle z_M \rangle)^2 \rangle\right)$, respectively.

We may now consider a smooth function $f(z)$ and consider its ensemble average and variance. Expanding $f(z)$ to second order, we obtain

$$\langle f(z) \rangle \simeq f(\langle z \rangle) + \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M} f''_{ij}(\langle z \rangle)\left\langle (z_i - \langle z_i \rangle)\left(z_j - \langle z_j \rangle\right)\right\rangle$$ (23)

[†]It is important to note that the subscripts here denote M different stochastic processes *and not* as in other contexts the number of one of the N realizations or trials.

and

$$\left\langle (f(z) - \langle f(z) \rangle)^2 \right\rangle \simeq \sum_{i=1}^{M} f_i'(\langle z \rangle) \sum_{j=1}^{M} f_j'(\langle z \rangle) \left\langle (z_i - \langle z_i \rangle) \left(z_j - \langle z_j \rangle \right) \right\rangle, \qquad (24)$$

where primes indicate partial differentiation with respect to the variable with the numbers in the lower indices.

The last equation shows the mechanism of error propagation: the statistical uncertainty of a single random variable z is often represented by a variance, the so-called error variance (see, e.g. Lenschow et al. (1994)). The error variance of a function $f(z)$ of z will then be determined by

$$\left\langle (f(z) - \langle f(z) \rangle)^2 \right\rangle \simeq f'^2(\langle z \rangle) \langle (z - \langle z \rangle)^2 \rangle. \qquad (25)$$

When there are more than one random variable the uncertainty of each variable z_i may again be determined by its error variance whereas the uncertainty of a function of these variables will include all the possible covariances of pairs of z_i and z_j. For example, in the case of two random variables we have

$$\begin{aligned} &\left\langle \left(f(z_1, z_2) - \langle f(z_1, z_2) \rangle \right)^2 \right\rangle \\ &\simeq f_1'^2(\langle z_1 \rangle, \langle z_2 \rangle) \langle (z_1 - \langle z_1 \rangle)^2 \rangle + f_2'^2(\langle z_1 \rangle, \langle z_2 \rangle) \langle (z_2 - \langle z_2 \rangle)^2 \rangle \\ &+ 2 f_1'(\langle z_1 \rangle, \langle z_2 \rangle) f_2'(\langle z_1 \rangle, \langle z_2 \rangle) \langle (z_1 - \langle z_1 \rangle) \langle (z_2 - \langle z_2 \rangle) \rangle. \end{aligned} \qquad (26)$$

A straightforward generalization of (24) to covariances between two functions $f(z)$ and $g(z)$ will also be needed in the following.

$$\begin{aligned} &\langle (f(z) - \langle f(z) \rangle) (g(z) - \langle g(z) \rangle) \rangle \\ &\simeq \sum_{i=1}^{M} f_i'(\langle z \rangle) \sum_{j=1}^{M} g_j'(\langle z \rangle) \left\langle (z_i - \langle z_i \rangle) \left(z_j - \langle z_j \rangle \right) \right\rangle. \end{aligned} \qquad (27)$$

Basic Statistics for Pairs

We assume that the pairs (x_i, y_i), $i = 1, 2, \ldots, N$ are independent and identically distributed. Without loss of generality we may also assume that they have zero ensemble means.

The *ensemble variances* are then

$$\langle x^2 \rangle \equiv \langle x_i^2 \rangle \qquad (28)$$

and

$$\langle y^2 \rangle \equiv \langle y_i^2 \rangle. \qquad (29)$$

We therefore have the following relations

$$\langle x_i x_j \rangle = \langle x^2 \rangle \delta_{ij}, \qquad (30)$$

$$\langle y_i y_j \rangle = \langle y^2 \rangle \delta_{ij}, \tag{31}$$

and

$$\langle x_i y_j \rangle = \sqrt{\langle x^2 \rangle \langle y^2 \rangle} \, \rho_0 \, \delta_{ij}, \tag{32}$$

where ρ_0 is the correlation coefficient.

The ensemble means can of course only be estimated by the sample means. We have

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{33}$$

and

$$\bar{y} = \frac{1}{N} \sum_{i=1}^{N} y_i, \tag{34}$$

and we assume that N is so large that we may consider $\bar{x} \simeq \bar{y} \simeq 0$ good approximations to the ensemble means.

The statistical uncertainties measured in terms of the variances of the samples means are

$$\left\langle (\bar{x} - \langle x \rangle)^2 \right\rangle = \frac{\langle x^2 \rangle}{N} \tag{35}$$

and

$$\left\langle (\bar{y} - \langle y \rangle)^2 \right\rangle = \frac{\langle y^2 \rangle}{N}. \tag{36}$$

The ensemble variances entering (35) and (36) are estimated by the *sample variances*

$$\overline{x^2} = \frac{1}{N} \sum_{i=1}^{N} x_i^2 \tag{37}$$

and

$$\overline{y^2} = \frac{1}{N} \sum_{i=1}^{N} y_i^2. \tag{38}$$

Similarly, the covariance of x_i and y_i and the correlation coefficient are approximated by

$$\overline{xy} = \frac{1}{N} \sum_{i=1}^{N} x_i y_i \tag{39}$$

and

$$\rho_0 = \frac{\langle xy \rangle}{\langle x^2 \rangle^{1/2} \langle y^2 \rangle^{1/2}} \simeq \frac{\overline{xy}}{\overline{x^2}^{1/2} \overline{y^2}^{1/2}} = \rho. \tag{40}$$

In order to determine the variances of $\overline{x^2}$, $\overline{y^2}$, and their covariance we assume that the probability density of (x_i, y_i) is joint Gaussian.

We then obtain

$$\left\langle \left(\overline{x^2} - \langle x^2 \rangle \right)^2 \right\rangle = 2\frac{\langle x^2 \rangle^2}{N}, \tag{41}$$

$$\left\langle \left(\overline{y^2} - \langle y^2 \rangle \right)^2 \right\rangle = 2\frac{\langle y^2 \rangle^2}{N}, \tag{42}$$

and

$$\left\langle \left(\overline{x^2} - \langle x^2 \rangle \right) \left(\overline{y^2} - \langle y^2 \rangle \right) \right\rangle = 2\frac{\langle x^2 \rangle \langle y^2 \rangle}{N}\rho_0^2. \tag{43}$$

Statistical Uncertainties of a and b

Now we have the tools for determining the statistical uncertainties of a and b given by (20) and (21).

In the present application the two variables x and y are highly correlated with a correlation coefficient close to unity. They are also very close to being identical which means that the absolute value of δ is much smaller than one. This implies that (20) in error analyses can safely be approximated by

$$a \simeq 1 - \delta. \tag{44}$$

Using (26) with

$$f(z_1, z_2) = 1 - \frac{1}{2} \ln\left(\frac{z_1}{z_2}\right) \tag{45}$$

with the random variables

$$\left\{ \begin{array}{c} z_1 \\ z_2 \end{array} \right\} = \left\{ \begin{array}{c} \overline{x^2} \\ \overline{y^2} \end{array} \right\}, \tag{46}$$

we have

$$f_1'(z_1, z_2) = -\frac{1}{2z_1} \tag{47}$$

and

$$f_2'(z_1, z_2) = \frac{1}{2z_2} \tag{48}$$

so that the variance of a becomes

$$\begin{aligned} \sigma^2\{a\} &\equiv \left\langle \left(f(z_1, z_2) - \langle f(z_1, z_2) \rangle \right)^2 \right\rangle \\ &= \frac{1}{4}\left\{ \frac{2}{N} - 2\frac{2\rho_0^2}{N} + \frac{2}{N} \right\} = \frac{1 - \rho_0^2}{N} \simeq 2\frac{1 - \rho_0}{N}. \end{aligned} \tag{49}$$

This equation has been confirmed by a Monte Carlo simulation as discussed in the appendix B.

The value of a is in our case very close to one. We can therefore simplify the expression for b when calculating the error variance of b as follows

$$b = \overline{y} - a\overline{x} \simeq \overline{y} - \overline{x}. \tag{50}$$

In this case

$$f(z_1, z_2) = z_2 - z_1 \tag{51}$$

with

$$\left\{ \begin{array}{c} z_1 \\ z_2 \end{array} \right\} = \left\{ \begin{array}{c} \overline{x} \\ \overline{y} \end{array} \right\} \tag{52}$$

so that

$$\sigma^2\{b\} = \frac{\langle x^2 \rangle + \langle y^2 \rangle - 2\langle xy \rangle}{N} \simeq \frac{\overline{x^2} + \overline{y^2} - 2\overline{xy}}{N}. \tag{53}$$

Finally, to be able to evaluate the statistical uncertainty of functions of a and b we need to determine the covariance

$$\mu\{a, b\} = \\ \left\langle \left(f(z_1, z_2, z_3, z_4) - \langle f(z_1, z_2, z_3, z_4) \rangle \right) \left(g(z_1, z_2, z_3, z_4) - \langle g(z_1, z_2, z_3, z_4) \rangle \right) \right\rangle, \tag{54}$$

where

$$\left\{ \begin{array}{c} z_1 \\ z_2 \\ z_3 \\ z_4 \end{array} \right\} = \left\{ \begin{array}{c} \overline{x^2} \\ \overline{y^2} \\ \overline{x} \\ \overline{y} \end{array} \right\}, \tag{55}$$

and where

$$f(z_1, z_2, z_3, z_4) = a = 1 - \frac{1}{2}\ln\left(\frac{z_1}{z_2}\right) \tag{56}$$

and

$$g(z_1, z_2, z_3, z_4) = b = z_4 - f(z_1, z_2, z_3, z_4)z_3 = z_4 - z_3 + \frac{z_3}{2}\ln\left(\frac{z_1}{z_2}\right). \tag{57}$$

Again, applying (27) we find that a and b are uncorrelated, that is

$$\mu\{a, b\} = 0. \tag{58}$$

Effective Number of Degrees of Freedom

Until now we have assumed that all the trials (x_i, y_i) are statistically independent. As a consequence the number of degrees of freedom has been set equal to the number N of trials. In general this assumption is not true and in particular in our case we must take into account that a pair of ten-minute averages of wind speed is not independent of the preceding pairs of ten-minute averages. The records we use have a "memory" which is conveniently described by the *integral time scale* $\mathscr{T}$, defined by the integral of the *autocorrelation function* $\rho(\tau)$ of the time series, in our case a stationary, continuous record of running 10-minute averages of a wind speed. We use the definition

$$\mathscr{T} = \int_0^\infty \rho(\tau)\,\mathrm{d}\tau, \tag{59}$$

where

$$\rho(\tau) = \langle [x(t) - \langle x \rangle][x(t+\tau) - \langle x \rangle] \rangle / \langle x^2 \rangle. \tag{60}$$

We see from these two equations that the shorter the memory, the faster $\rho(\tau)$ decreases with the separation time τ, and the smaller the integral scale.

We consider the stationary and continuous time series $x(t)$ with the time average (33) with the summation index i representing the order of the observation time. As justified later, we deal with a separation Δt between observations so small that summations can be replaced by integrations. For completeness, the more general case where Δt cannot be considered small is discussed in appendix C. Thus we have

$$
\begin{aligned}
\bar{x} &= \frac{1}{N}\sum_{i=1}^{N} x_i \equiv \frac{1}{N}\sum_{i=1}^{N} x(i\Delta t)\underbrace{\Delta i}_{=1} \\
&= \frac{1}{N}\sum_{i=1}^{N} x(i\Delta t)\frac{\Delta i}{\Delta t}\Delta t \simeq \frac{1}{T}\int_0^T x(t)\,\mathrm{d}t,
\end{aligned}
\tag{61}
$$

where $T = N\Delta t$ is the observation time.

With this simplification the ensemble variance of $\bar{x}$—the error variance—becomes

$$
\begin{aligned}
\sigma^2\{x\} &\equiv \left\langle (\bar{x} - \langle x \rangle)^2 \right\rangle = \frac{\langle x^2 \rangle}{T}\int_0^T \mathrm{d}t'\frac{1}{T}\int_0^T \mathrm{d}t''\rho(t''-t') \\
&= \frac{2\langle x^2 \rangle}{T}\int_0^T \left\{1 - \frac{\tau}{T}\right\}\rho(\tau)\,\mathrm{d}\tau.
\end{aligned}
\tag{62}
$$

A comparison between (62) and (35) leads us to define an *effective number of degrees of freedom* by

$$\frac{1}{N_{\text{eff}}} = \frac{\sigma^2\{x\}}{\langle x^2 \rangle} = \frac{2}{T}\int_0^T \left\{1 - \frac{\tau}{T}\right\}\rho(\tau)\,\mathrm{d}\tau. \tag{63}$$

To determine this number we must know the autocorrelation function $\rho(\tau)$. This is obtained by the model

$$\rho(\tau) = \mathrm{e}^{-|\tau|/\mathscr{T}}. \tag{64}$$

The power spectrum then becomes

$$S(f) = \int\limits_{-\infty}^{\infty} \rho(\tau)\,\mathrm{e}^{-2\pi i f \tau}\mathrm{d}\tau = \frac{2\mathscr{T}}{1 + (2\pi\mathscr{T}f)^2}. \tag{65}$$

We have analyzed an almost 17-year long time series of 10-minute averaged wind speed signal measured at the top of a 40 m mast in Tystofte in southern Zealand (Denmark). Details about these data can be found in Kristensen et al. (1999).

Using a standard FFT routine (Fast Fourier Transform), we calculated $S(f)$. This spectrum is shown in Fig. 5 in an area-conserving, log-linear plot. We fitted (65) to this spectrum and found $\mathscr{T} \simeq 20.2$ h.
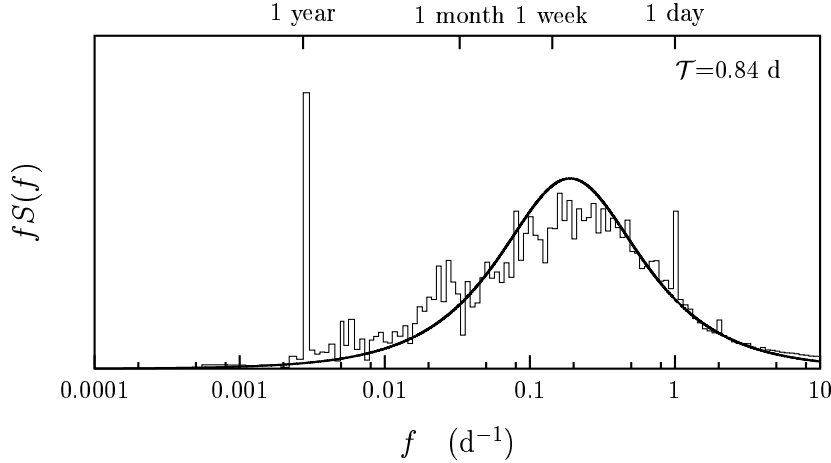

Figure 5. Power spectrum of the horizontal wind speed at Tystofte in southern Zealand. The data consisted of almost 17 years of ten-minute averages and were obtained by a cup anemometer at the height of 40 m over rural terrain. The annual and diurnal periods are quite pronounced although the corresponding peaks comprise only about 5% of the total variance. The solid line is a fit of the form (65) to the spectrum without these two peaks.

Since the total number of observations N is given by $T/\Delta t$, where $\Delta t = 10$ min is the time between observations, the expression for the effective number of degrees of freedom (63) can now be evaluated as a function of N with $q = \mathscr{T}/\Delta t$ as parameter:

$$\frac{N_{\mathrm{eff}}}{N} = \frac{1}{2q\left\{1 - \frac{q}{N}\left(1 - \mathrm{e}^{-N/q}\right)\right\}}. \tag{66}$$

In our case where we are concerned with 10-minute averages of wind speed the value of the parameter q is fixed and equal to about $20.2\ h/(1/6\ h) \simeq 121$. There are two limiting cases:

$$\frac{N_{\mathrm{eff}}}{N} \approx \begin{cases} \frac{1}{N}\left\{1 + \frac{N}{3q}\right\} & , \quad N \ll q \\[2mm] \frac{1}{2q}\left\{1 + \frac{q}{N}\right\} & , \quad N \gg q. \end{cases} \tag{67}$$
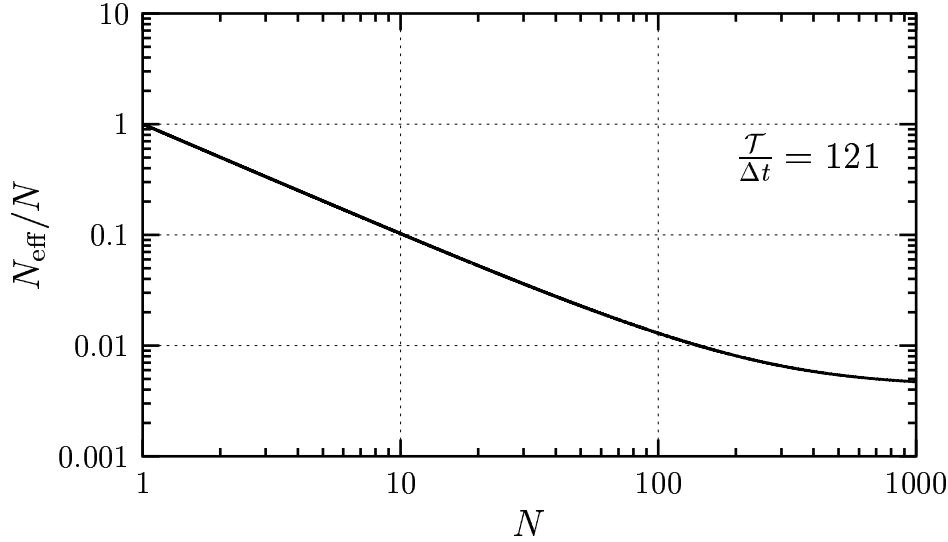
Figure 6. The ratio N_{eff}/N as a function of N for $q = \mathcal{T}/\Delta t = 121$.

The ratio N_{eff}/N as a function of N with $q = 121$ is shown in Fig. 6.

The result for $N \gg q$ in (67) is actually more general as can be seen by inspecting (62) for $T \to \infty$. In view of the definition (59) we get

$$\sigma^2\{x\} \simeq 2\langle x^2 \rangle \frac{\mathcal{T}}{T}, \tag{68}$$

from which the second line of (67) follows directly. This general expression is equivalent to

$$\sigma^2\{x\} = \langle x^2 \rangle \frac{S(0)}{T}, \tag{69}$$

which can easily be shown by means of (59) and the first part of (65).

Since $\mathcal{T}$ represents a time interval within which the correlation of the values of the time series at two times cannot be neglected, we would expect that N_{eff} would be equal to N if just $\Delta t \gg \mathcal{T}$. This is in fact the case, but in order to deal with this situation the more general approach in appendix C must be applied.

So far we have discussed the effective degrees of freedom pertaining the ensemble variance of the sample mean $\bar{x}$. When it comes to the ensemble variance of the sample variance the equation for N_{eff} is no longer valid. However, as shown by Lenschow et al. (1994), the error variance in this case can be determined if the skewness and the kurtosis of the time series $x(t)$ are known. The same data from which the power spectrum Fig. 5 was calculated showed that skewness and kurtosis are 0.6 and 3.6. The theory by Lenschow et al. (1994) then predicts that the effective number of degrees of freedom is reduced to about $0.75 \times N_{\text{eff}}$ for the variance of the sample variance.

We may test in another way that there is a reduction in degrees of freedom when the samples are correlated, namely by calculating directly what in the turbulence community could be called the "one-step structure function"

$$d^2 = \frac{1}{N-1} \sum_{i=2}^{N} (x_i - x_{i-1})^2. \tag{70}$$

The ensemble mean of this quantity becomes

$$
\begin{aligned}
\langle d^2 \rangle &= \frac{1}{N-1} \sum_{i=2}^{N} \langle (x_i - x_{i-1})^2 \rangle \\
&= \frac{1}{N-1} \sum_{i=2}^{N} \langle x_i^2 \rangle + \frac{1}{N-1} \sum_{i=2}^{N} \langle x_{i-1}^2 \rangle - \frac{2}{N-1} \sum_{i=2}^{N} \langle x_i x_{i-1} \rangle \\
&= 2 \langle x^2 \rangle \{ 1 - \rho_1 \},
\end{aligned}
\tag{71}
$$

where ρ_1 is the correlation coefficient between two successive observations.

We see that if the observations are independent such that the correlation coefficient between neighboring observations is zero then the one-step structure function is exactly twice the variance $\langle x^2 \rangle$.

Thus, if we "dilute" the data used for calculating the spectrum in Fig. 5 so that we only use $N_{\text{eff}} \leq N$ equidistantly spaced observations, then we would expect that $d^2/2$ would be almost equal to $\overline{x^2}$. Pretending that we have no prior knowledge of the integral scale $\mathcal{T}$, we compute $\overline{x^2}$ and d^2 directly from the observations with a number of assumptions about the ratio N/N_{eff}. The result is given in Table 2.

Table 2. The sample variance and the one-step structure function for an increasing degree N/N_{eff} of dilution of the time series $x(t)$ of velocities measured at Tystofte.

N/N_{eff}	$\overline{x^2}$	$d^2/2$
	$(\text{m/s})^2$	$(\text{m/s})^2$
1	9.48	0.18
5	9.48	0.56
10	9.49	0.91
50	9.47	3.29
100	9.55	5.11
500	9.45	8.10

We see that when we set the "dilution factor" N/N_{eff} equal 500, the one-step structure function divided by two becomes almost equal to the variance. This means that $\mathcal{T}$ determined from the spectral fit provides an estimate of N_{eff} which, within a factor of two, is consistent with that obtained by studying the ratio $\overline{x^2}/d^2$ as a function of the dilution of the data.

Before we discuss the analysis of an actual data record we must consider the problem that the time series we analyze will be intermittent because we impose conditions on both the lowest wind speed and the direction sector for the inclusion of the measured wind speed values. Figure 10 shows a typical example. There are two opposing effects of the intermittency on the effective number of degrees of freedom N_{eff}. The most obvious is that the number of observations used is reduced compared the the total number. This in itself will reduce N_{eff}. The other effect is that some of the observations will be spaced more in time, thus enhancing N_{eff} because their mutual statistical dependence is reduced. This, rather complicated, problem is discussed in appendix D. The result is that when

the total observation time T is much larger than the integral time scale $\mathscr{T}$ then the error variance, i.e. the ensemble variance of the intermittently sampled mean of $x(t)$, becomes

$$\sigma^2\{x\} = 2\langle x^2 \rangle \frac{\mathscr{T}}{T}\left\{1 + \frac{\sigma_\chi^2/\langle\chi\rangle^2}{1 + \frac{\eta\mathscr{T}}{\sigma_\chi^2(1-2\sigma_\chi^2)}}\right\}, \tag{72}$$

where $\langle\chi\rangle$ and σ_χ^2 are the ensemble mean and variance of the time series $\chi(t)$ which is one when the signal $x(t)$ is fulfilling the inclusion criteria and zero otherwise, and where η is the average rate of change from $\chi(t) = 1$ to $\chi(t) = 0$.

By comparing (72) with (68) we see that $\sigma^2\{x\}$ is the usual error variance for the entire time seriens with a correction factor. Instead of the lower equation in (67) which essentially states $N_{\text{eff}}/N = \Delta t/(2\mathscr{T})$ we have

$$\frac{N_{\text{eff}}}{N} = \frac{\Delta t}{2\mathscr{T}} \times \left\{1 + \frac{\sigma_\chi^2/\langle\chi\rangle^2}{1 + \frac{\eta\mathscr{T}}{\sigma_\chi^2(1-2\sigma_\chi^2)}}\right\}^{-1}. \tag{73}$$

4   Data Analysis

The data-analysis process is here illustrated in a case where all the anemometers are Risø anemometers of the type shown in Fig. 1.

It is a general experience with cup anemometers that below a certain, anemometer-dependent wind speed the calibration ceases to be linear. In fact, the minimum wind speed at which an anemometer starts rotation need not be the same at which it stops when the wind speed is decreasing. For most cup anemometers the wind speed below which these complications are important is about 1 m/s. Here we choose a minimum 3 m/s wind speed. This decision is also influenced by the fact that in the wind tunnels where reference calibrations are carried out, the lowest wind speed is set to about 4 m/s. We have also chosen a maximum wind speed of 16 m/s, equal to that used in wind-tunnel calibration. This last choice is of little consequence because such high wind speeds are quite uncommon at the Risø calibration site (Figs. 2 and 3).

In the period from the morning of June 13 through the morning of July 11, 2000 we had a rather long period with wind coming from west, over water. The 10 wind-speed records are shown in Fig. 7.

The anemometer P300 in position Ref 1 was used as reference. This anemometer has been wind-tunnel calibrated by WINDTEST KWK GmbH (KWK in the following) in Hamburg. The following linear relation between the wind speed U and the pulse frequency f_0 in Hz was obtained:

$$U = A_0 \times f_0 + B_0, \tag{74}$$

where $A_0 = 0.61602$ m and $B_0 = 0.255$ m/s.

The cup anemometers P470...P477 in the test positions were also first calibrated by KWK and this gave us the opportunity to compare these calibrations with those obtained
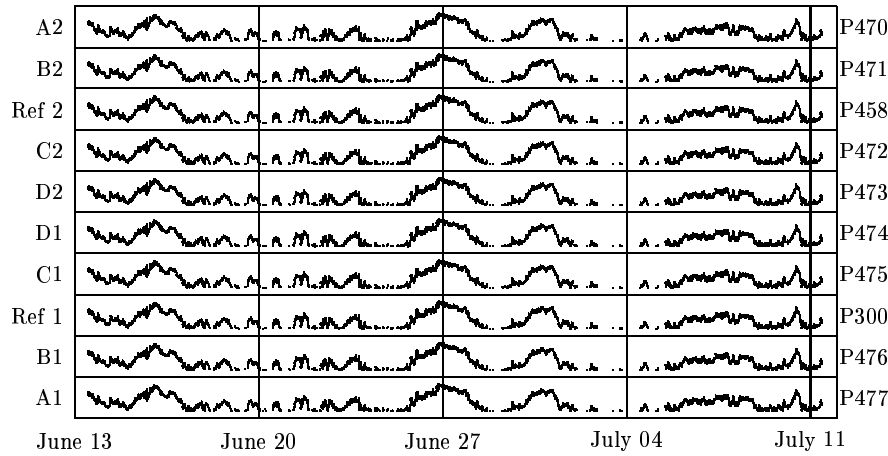
Figure 7. Records of 10-minute averages of the raw anemometer signals from the calibration period from June 13, 2000, 10:45 until July 11, 2000, 10:05. Signal values corresponding to 3 m/s or less are excluded. This corresponds to 74% of the data. The range in each of the 10 records is 0–1024. To the left are shown the position names and to the right the anemometer code names.

from the boom measurements (Fig. 7) by comparing their signals with the signal from the reference anemometer P300.

The technique was to fit the raw datalogger signals of the eight test anemometers to the raw reference signal according to the method described in section 3. Converting the raw signals to mean frequencies by multiplying by 32 and dividing by 600 s (see section 2), we determine a and b in the equation

$$f_0 = a \times f + b, \tag{75}$$

where the frequency f pertains to the reference anemometer.

The calibration expression for the test anemometer is

$$U = A \times f + B \tag{76}$$

and, since the test anemometer and the reference anemometer supposedly have been exposed to the same wind history, we have

$$A \times f + B \equiv A_0 \times f_0 + B_0 = A_0(a \times f + b) + B_0, \tag{77}$$

which implies

$$A = aA_0 \tag{78}$$

and

$$B = B_0 + bA_0. \tag{79}$$

First we selected data in 30° sectors around -60°, -30°, 0°, 30°, and 60°. The result is shown in Fig. 8.

By inspecting Fig. 8, we see that the boom calibrations are in reasonable agreement with the wind-tunnel calibrations when the wind comes from west in a broad 150° sector.
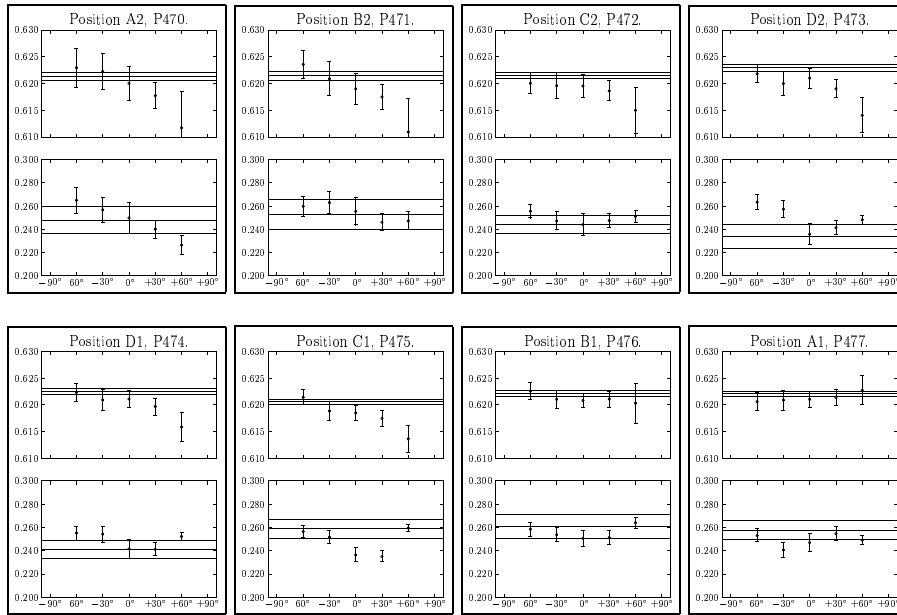
Figure 8. Comparison between wind-tunnel calibrations and boom calibrations. These wind-speed data are divided into five different 30° direction sectors. The anemometer P300 in position Ref 1 *is used as reference for* P470...P477. *There is one frame for each of these anemometers where the upper subframe pertains to the slope A and the lower to the offset B. The eight anemometers,* P470...P477, *have been wind-tunnel calibrated on the same day, June 14, 2000, and the result is shown as three horizontal lines indicating the mean in the middle while the other two are the mean ± one standard deviation (68% confidence limits). The results from the boom calibrations with* P300 *as reference are shown as points with standard deviations for each direction sector. This reference anemometer has been calibrated in the same wind tunnel, but on March 1, 2000.*

However, the values of *A* in particular seem to fall below the wind-tunnel values. This tendency could be caused by the fact that the eight test anemometers are wind-tunnel calibrated on another date than the reference anemometer. We have tested if this should be the case by using the anemometer P475 in position C1 as reference. The result is shown in Fig. 9. The tendency for *A* to be a little too small seems to have disappeared (except for P300 now considered a test anemometer), but otherwise the degree of consistency between boom calibration and wind-tunnel calibration appears to be the same as when P300 is the reference.

Being slightly cautious, we will use data from the sector $0° \pm 45°$ in the following. Figure 10 shows how much of the data is used in the intercalibration in one particular case. It also gives an indication of the distribution and duration of periods where $\chi(t)$ is one. The total time is 51% so that $\langle \chi \rangle = 0.51$ and $\sigma_\chi^2 = 0.25$. The average rate of change from $\chi(t) = 1$ to $\chi(t) = 0$ is $\eta = 2.4 \, \mathrm{d}^{-1}$. As shown in the previous section, these data are important for determining the statistical uncertainty of the calibration parameters.

At this point it seems natural to test whether the flow around a particular cup anemometer is influenced by the neighbor instruments on the boom from which it is separated by the distance 0.75 m. If the flow must be considered disturbed we would of course expect the disturbance to be most pronounced when the wind direction is far from being perpendicular to the boom. To investigate this problem we compared a special set of data, recorded in the period from August 8, 2000 until August 28, 2000. In this period the positions A2, B2, C2, and D2 were unoccupied whereas Risø anemometers were occupying for all the other were 6 positions. Just as in the first measuring period in June-July, 2000, P458
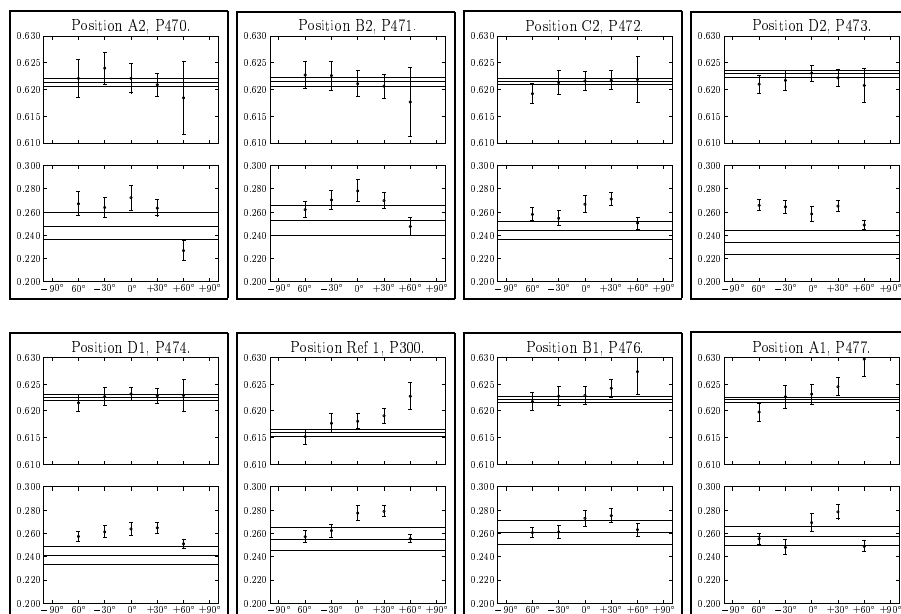
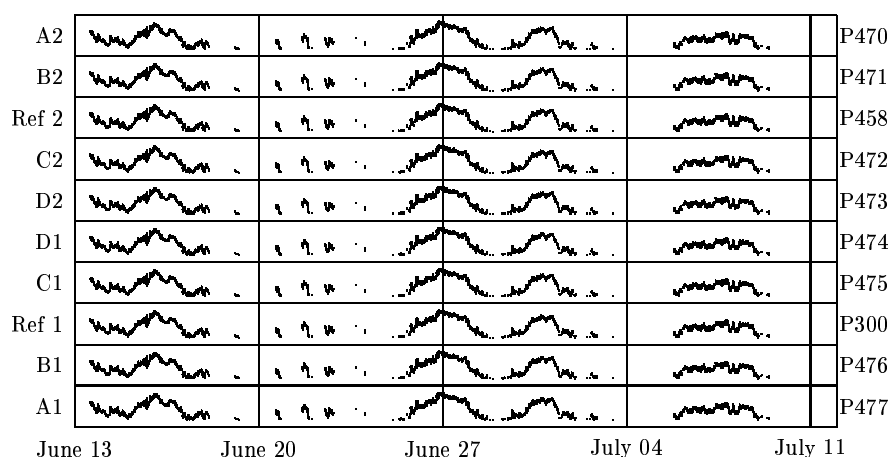*Figure 9. Same as Fig. 8, but with* P475 *in position* C1 *as reference.*

Figure 10. Records of 10-minute averages of the raw anemometer signals in the direction sector $0° \pm 45°$ *from the calibration period from June 13, 2000, 10:45 until July 11, 2000, 10:05. Signal values corresponding to 3 m/s or less are excluded. Now 51% of the data fulfill the speed and direction selection criteria.*

and P300 occupied positions Ref 2 and Ref 1, respectively, but now the distance from P458 to the nearest anemometer in position D1 was 2.25 m (see Fig. 3). The result of the comparison between P458 and the reference anemometer P300 is shown in Fig. 11, where both the first data set with all positions occupied and the last set of data from August are used. If there were no influence on the flow from the disturbance from neighbor anemometers, both constants A and B should be the same in the two instruments. These constants are shown with their 68% confidence limits and, apparently, the values of A are more in agreement than the values of B. We see, however, that the dependence of the disagreement on wind direction is not pronounced. Using data from a 90° direction sector centered around 0°, we find $A = 0.6196 \pm 0.0026$ m and $B = 0.2442 \pm 0.0104$ m/s for the June data and $A = 0.6175 \pm 0.0040$ m and $B = 0.2703 \pm 0.0106$ m/s for those of Au-

gust. Applying the so-called Z-test for two sample means to be the same (Kanji 1999) at $\alpha = 5\%$, we find that $Z = -0.45$ for A and $Z = 2.5$ for B. The $\alpha = 0.05$ limit for rejection is $Z = 1.96$ so we conclude that the calibration offset B is influenced by flow distortion from neighbor instruments whereas we should not reject that gain A is the same whether it has a close neighbor instrument or not. When the wind speed is more than a few meters per second the accuracy of the gain is much more important than the offset for reliable wind-speed measurements. In other words, the systematic error on B, which may amount to about 0.02 m/s, can in most situations be neglected.
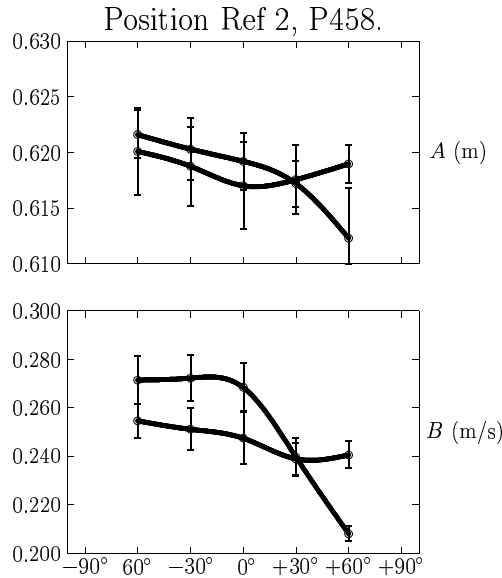

Figure 11. Comparison between calibrations of cup anemometer P458 *in position Ref 2 when all boom positions are occupied (period June 2000) and when positions* A2, B2, C2, *and* D2 *are empty, i.e. when* P458 *has no close neighbors (August 2000). In the upper frame the values of A from the June period are larger than those pertaining to the August period, except when the center direction is* $+60°$. *For the values of B it is just the opposite. The 68% confidence limits are shown at each point.*

5 Conclusion

It has now been demonstrated how it is possible to simultaneously calibrate several cup anemometers in open air with a single anemometer as the reference. It requires that all the anemometers, including the reference, are exposed to the same wind flow. To guarantee this, care must be taken that the anemometers do not interfere with each other and that, in general, the upstream conditions should look the same for all the anemometers. The Risø calibration facility is, as described in section 2, a boom with 10 anemometer mounts, erected parallel to the west coast of Roskilde Fjord and with a several-kilometer fetch of water surface in a broad direction sector towards west.

Compared to wind-tunnel calibration, there are advantages and disadvantages.

Obviously it is advantageous to be able to calibrate many anemometers simultaneously at low labor cost. In particular when the anemometers are later to be used for accurate comparison of wind fields at different locations, i.e. along a vertical mast in cases where reliable wind profiles are required. Another bonus is that new cup anemometers can be op-

erated in the field before they are deployed for real field measurements: a cup anemometer needs some time, typically one month, to be "broken in" if a new rotor with new bearings has been installed.

One disadvantage is that the method obviously does not provide an absolute calibration with reference to a certified wind tunnel. Such a calibration is often required by wind turbine manufactures and owners for them to be able to settle if a particular wind turbine has produced the expected electric energy. Another disadvantage is that it may take very long time to obtain a calibration because there are limits to both direction and magnitude of the free wind.

In the analysis of data from the test facility we found it natural to use orthogonal mean square fitting instead of the usual linear fitting with one independent and one dependent variable. This is so because the signals from all the instruments, including reference anemometers, are almost equal. Section 3 and the appendices B, C, and D contain a rather detailed discussion of the fitting procedure and the statistical significance of the results. One item of particular importance is that, in contrast to measurements in wind tunnels, consecutive sets of data cannot be considered statistically independent. In fact, the time scale (or memory) of ten-minute wind speed averages is shown to be about 20 hours. In principle this means that data sets must be separated in time more than about 40 hours to be considered statistically independent.

We have tested whether the flow around a particular anemometer is disturbed by the presence of the neighbor anemometers and we found that the calibration gain A appears unaffected whereas the offset B seems influenced in a systematic way. However, when the wind speed is more than a few meters per second the offset, being itself about 0.2 m/s, is of little importance.

The question is if the field calibration is really preferable to wind tunnel calibration. To calibrate a cup anemometer will in the last case typically take about half an hour. There might be a slight overhead for setting up a calibration stand, but altogether a set of 10 anemometers can be calibrated in the course of one day. It can also be argued that a generally approved calibration can be obtained only by use of a certified wind tunnel. The problem with this is of course that there are several certified wind tunnels and that they do not always give the same calibration for the same anemometer.

This is probably so because the reference wind speed in a tunnel is determined by measuring the very small pressure differences (about 0.1 HPa or less) from a Pitot tube. The pressure detector is temperature sensitive and outmost care must be taken by monitoring the temperature of the air in the wind tunnel. Even then there seems to be problems and to illustrate this point we have had the eight cup anemometers calibrated not only in the KWK wind tunnel, but also in the wind tunnel of Svend Ole Hansen Aps (SOH). The first calibration took place on June 14, 2000, the second on August 23, 2000. The calibration results are summarized in Table 3. We see that for all eight anemometers the measured gains A are larger in the SOH tunnel whereas the opposite is the case for the offsets B. In fact, the differences of the last are all between 0.1 m/s and 0.2 m/s. Judging by the confidence limits, the differences do not seem to be within the statistical variability.

To illustrate how much the difference in the two calibrations influence the actual measured wind speeds, we have plotted the velocity difference with 68% confidence limits over a range of about 0 to 20 m/s for one of the cup anemometers, P475. This is shown in Fig. 12. We see that the calibrations agree within ±0.05 m/s in a wind speed range from about 7 m/s to about 12 m/s. This relatively poor agreement in calibrations seems somewhat disappointing in view of the high quality of the anemometers in terms of long-term stability of the calibration.

Perhaps the best solution to obtain reliable calibrations of cup anemometers is to modify

Table 3. Comparison between wind tunnel calibrations at KWK *and* SOH*. The gains A and the offsets B are given with 68% confidence limits.*

	KWK		SOH	
	A (m)	*B* (m/s)	*A* (m)	*B* (m/s)
P470	0.6213 ± 0.0007	0.247 ± 0.012	0.6299 ± 0.0007	0.118 ± 0.011
P471	0.6213 ± 0.0008	0.253 ± 0.013	0.6299 ± 0.0007	0.109 ± 0.010
P472	0.6214 ± 0.0005	0.244 ± 0.008	0.6289 ± 0.0007	0.152 ± 0.011
P473	0.6230 ± 0.0007	0.231 ± 0.010	0.6303 ± 0.0011	0.121 ± 0.016
P474	0.6223 ± 0.0005	0.248 ± 0.008	0.6284 ± 0.0009	0.144 ± 0.014
P475	0.6203 ± 0.0005	0.263 ± 0.008	0.6282 ± 0.0011	0.138 ± 0.016
P476	0.6218 ± 0.0005	0.265 ± 0.010	0.6298 ± 0.0012	0.148 ± 0.017
P477	0.6218 ± 0.0005	0.262 ± 0.008	0.6288 ± 0.0011	0.157 ± 0.016

the wind-tunnel calibration procedure by using a "standard" cup anemometer, rather than a Pitot-tube, as reference anemometer.
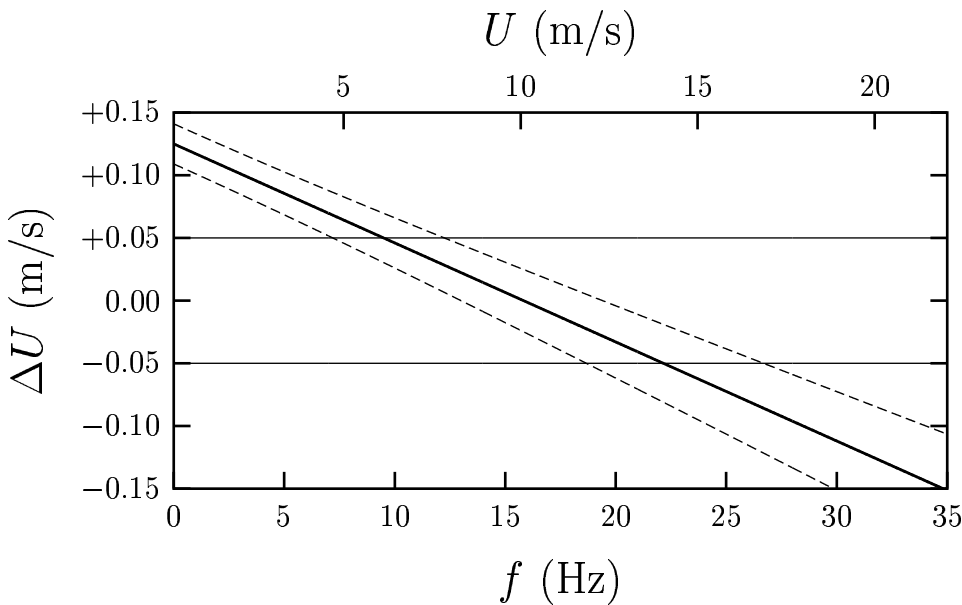
Figure 12. Comparison between calibrations of cup anemometer P475 *in the* KWK *wind tunnel and the* SOH *wind tunnel. The solid, thick line is the wind speed calculated with the* KWK *calibration parameters minus the wind speed calculated with the* SOH *calibration as function of frequency f. The corresponding wind speed, using the average of the two calibrations is shown along the upper abscissa. The dashed lines indicate the 68% confidence lines.*

A   Regression with Curvature

We want to investigate the orthogonal regression to a curve which is "almost" straight.

We use (2) in the generalized form

$$r = r_0 + \theta\, t(\theta). \tag{A1}$$

The parameter θ is the length of the chord from a reference point $r_0 = x_0 i + y_0 j$ on the curve to the curve point $r(\theta) = xi + yj$.

The expression (5) for the unit vector t may still be used if we consider α a function of θ.

From

$$\left\{ \begin{array}{c} x \\ y \end{array} \right\} = \left\{ \begin{array}{c} x_0 + \theta \cos(\alpha(\theta)) \\ y_0 + \theta \sin(\alpha(\theta)) \end{array} \right\} \tag{A2}$$

we obtain

$$\theta^2 = (x - x_0)^2 + (y - y_0)^2. \tag{A3}$$

We consider only small curvatures and assume the approximation

$$\cos(\alpha(\theta)) \simeq \cos\alpha_0 - \alpha_0'\theta\, \sin\alpha_0$$
$$\sin(\alpha(\theta)) \simeq \sin\alpha_0 + \alpha_0'\theta\, \cos\alpha_0, \tag{A4}$$

where $\alpha_0 = \alpha(0)$ is the direction of the tangent in the reference point r_0 and $\alpha_0' = \alpha'(0)$ its derivative.

This implies that (A2) can be written

$$x - x_0 = \theta\cos\alpha_0 - \alpha_0'\theta^2 \sin\alpha_0$$
$$y - y_0 = \theta\sin\alpha_0 + \alpha_0'\theta^2 \cos\alpha_0, \tag{A5}$$

or, by using (A3) to eliminate θ^2,

$$(x - x_0) + \alpha_0'\left[(x - x_0)^2 + (y - y_0)^2\right]\sin\alpha_0 = \theta\cos\alpha_0$$
$$(y - y_0) - \alpha_0'\left[(x - x_0)^2 + (y - y_0)^2\right]\cos\alpha_0 = \theta\sin\alpha_0. \tag{A6}$$

Multiplying the first equation by $\sin\alpha_0$ and the second by $\cos\alpha_0$ and subtracting, we get the following equation for the curve

$$\left(x - x_0 + \frac{\sin\alpha_0}{2\alpha_0'}\right)^2 + \left(y - y_0 - \frac{\cos\alpha_0}{2\alpha_0'}\right)^2 = \left(\frac{1}{2\alpha_0'}\right)^2, \tag{A7}$$

which is the equation for a circle with radius equal to $1/|2\alpha_0'|$ and center in the point $(x_0 - \sin\alpha_0/[2\alpha_0'], y_0 + \cos\alpha_0/[2\alpha_0'])$.

It is now very simple to determine the perpendicular (signed) distance d_i from a point with the coordinates (x_i, y_i) to the curve. It is simply the distance to the center of the circle minus the radius.

$$
\begin{aligned}
d_i &= \sqrt{\left(x_i - x_0 + \frac{\sin\alpha_0}{2\alpha_0'}\right)^2 + \left(y_i - y_0 - \frac{\cos\alpha_0}{2\alpha_0'}\right)^2} - \frac{1}{2|\alpha_0'|} \\
&\simeq \frac{1}{2|\alpha_0'|}\Big\{1 + 2\alpha_0'\left[(x_i - x_0)\sin\alpha_0 - (y_i - y_0)\cos\alpha_0\right] \\
&\qquad\quad + 2\alpha_0'^2\left[(x_i - x_0)^2 + (y_i - y_0)^2\right] \\
&\qquad\quad - 2\alpha_0'^2\left[(x_i - x_0)\sin\alpha_0 - (y_i - y_0)\cos\alpha_0\right]^2 - 1\Big\} \\
&= \frac{\alpha_0'}{|\alpha_0'|}\Big[(x_i - x_0)\sin\alpha_0 - (y_i - y_0)\cos\alpha_0 \\
&\qquad\quad + \alpha_0'\left\{(x_i - x_0)\cos\alpha_0 + (y_i - y_0)\sin\alpha_0\right\}^2\Big].
\end{aligned} \tag{A8}
$$

For N points (x_i, y_i), $i = 1, 2, \dots, N$ we calculate the distances d_i to the circle and minimize

$$
\phi = \frac{1}{N}\sum_{i=1}^{N} d_i^2 \tag{A9}
$$

to obtain the best fit to the circle (A7) in terms of the parameters x_0, y_0, α_0, and α_0' of which only three are independent.

At this point we assume that we can determine α_0' with sufficient accuracy by first finding x_0, y_0, and α_0 under the assumption that the line is straight and then calculating α_0' by solving $\partial\phi/\partial\alpha_0' = 0$ with respect to α_0', with x_0, y_0, and α_0 assumed known. Here we have the freedom to set $(x_0, y_0) = (0,0)$ and consequently obtain the solution

$$
\alpha_0' = -\frac{\overline{(x\sin\alpha_0 - y\cos\alpha_0)\left(x\cos\alpha_0 + y\sin\alpha_0\right)^2}}{\overline{\left(x\cos\alpha_0 + y\sin\alpha_0\right)^4}}. \tag{A10}
$$

We can now study how the slope $a = a(\theta) = \tan(\alpha)$ varies. We have

$$
a = \tan\left(\alpha_0 + \alpha_0'\theta\right) \simeq \tan\alpha_0 + \left(1 + \tan^2\alpha_0\right)\alpha_0'\theta. \tag{A11}
$$

To compare the *change* of a with θ with a itself we must select a range $\Delta\theta$ of θ. We have chosen the square root of sample variance of θ as this range:

$$
\Delta\theta = \sqrt{\overline{\theta^2}} = \sqrt{\overline{x^2 + y^2}}. \tag{A12}
$$

The corresponding change in a is then

$$
\Delta a \equiv a - \tan\alpha_0 = \left(1 + \tan^2\alpha_0\right)\alpha_0'\Delta\theta \simeq (1 + a^2)\alpha_0'\Delta\theta. \tag{A13}
$$

B   Monte Carlo Simulations

We have made Monte Carlo simulations in order to establish a numerical verification of (49), and also to study the quality of the approximation behind this formula, for a selected sequence of ρ_0-values.

To carry out this task we needed a procedure for producing standardized normal variates, assuming that a supply of independent, uniformly distributed (pseudo) random numbers $\xi_i \in U(0,1)$ are available from the computer. Such a method was designed by Box & Muller (1958) who proposed to generate pairs of independent values by the recipe

$$x = (-2\ln\xi_1)^{1/2}\cos(2\pi\xi_2) \tag{B1}$$

and

$$y = (-2\ln\xi_1)^{1/2}\sin(2\pi\xi_2). \tag{B2}$$

It is easy to verify that

$$x \in N(0,1), \tag{B3}$$

$$y \in N(0,1), \tag{B4}$$

and that

$$\text{Cov}[x,y] = 0. \tag{B5}$$

First we note that (B1) produces positive and negative numbers with equal probabilities. We may therefore consider positive values only and thus replace (B1) by

$$x = (-2\ln\xi_1)^{1/2}\cos(\tfrac{1}{2}\pi\xi_2) = uv. \tag{B6}$$

Let f, g, h be density functions for x, u, v, respectively, and F, G corresponding distribution functions. Then we have

$$F(x) = \int_0^1 h(v)\,G\!\left(\frac{x}{v}\right)\mathrm{d}v \tag{B7}$$

or

$$f(x) = \int_0^1 h(v)\,g\!\left(\frac{x}{v}\right)\frac{\mathrm{d}v}{v}. \tag{B8}$$

Now

$$g(u) = \left|\frac{\mathrm{d}\xi_1}{\mathrm{d}u}\right| = u\,\mathrm{e}^{-\frac{1}{2}u^2} \tag{B9}$$

and

$$h(v) = \left|\frac{\mathrm{d}\xi_2}{\mathrm{d}v}\right| = \frac{2}{\pi}\frac{1}{\sqrt{1-v^2}} \tag{B10}$$

Thus

$$f(x) = \frac{2}{\pi}\int_0^1 \frac{x}{v^2\sqrt{1-v^2}}\exp\!\left(-\frac{1}{2}\frac{x^2}{v^2}\right)\mathrm{d}v. \tag{B11}$$

This integral can be evaluated by substituting $v = x/\sqrt{x^2+u^2}$. The result is

$$f(x) = 2\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}x^2} \tag{B12}$$

which is the double of the normal density, as it should be. Now (B4) follows immediately from (B3), and (B5) follows by symmetry reasons □.

What we need next is a method for generating correlated pairs of normal deviates, sampled for a given correlation coefficient ρ_0. We first generate (x,y) as in (B1) and (B2). Then we take

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} c & \sqrt{1-c^2} \\ \sqrt{1-c^2} & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \tag{B13}$$

where c is a parameter in the interval $[-1,1]$. From (B13) we find the covariance matrix

$$D = \begin{pmatrix} c & \sqrt{1-c^2} \\ \sqrt{1-c^2} & c \end{pmatrix}^2 = \begin{pmatrix} 1 & 2c\sqrt{1-c^2} \\ 2c\sqrt{1-c^2} & 1 \end{pmatrix}. \tag{B14}$$

We must require

$$2c\sqrt{1-c^2} = \rho_0 \tag{B15}$$

or

$$c^2 = \frac{1 - \sqrt{1 - \rho_0^2}}{2}. \tag{B16}$$

Now we are in a position to simulate pairs (x_i, y_i), $i = 1,\dots,N$ such that $x_i \in N(0,1)$, $y_i \in N(0,1)$, and $\text{Cov}[x_i, y_i] = \rho_0$. Our choice $V[x_i] = V[y_i] = 1$ is natural in view of our knowledge that $V[x_i] \simeq V[y_i]$. Next we compute

$$s_x^2 = \overline{x_i^2}, \qquad s_y^2 = \overline{y_i^2}, \qquad s_{xy} = \overline{x_i y_i} \tag{B17}$$

and, according to (17) and (18),

$$\delta = \frac{1}{2} \ln\left(\frac{s_x^2}{s_y^2}\right), \tag{B18}$$

$$\varepsilon = \frac{1}{2} \ln\left(\frac{s_x^2 s_y^2}{s_{xy}^2}\right), \tag{B19}$$

and finally a by (20). We repeat the simulation for $j = 1,\dots,M$, each time recording $a = a_j$. In this way we can compute the sample variance $s^2(a)$ of a. Finally we compare the result with (49).

We used a fixed number of $M = 100$ repetitions, and with this we ran simulations for $N = 10, 100, 1000$, and 10000. For each N we took $\rho_0 = 0.9, 0.99$, and 0.999. The results are shown in the following tables:

	ρ_0	$2(1-\rho_0)/N$	$s^2(a)$
	0.9	$2.00.10^{-2}$	$2.74.10^{-2}$
$N = 10$	0.99	$2.00.10^{-3}$	$3.23.10^{-3}$
	0.999	$2.00.10^{-4}$	$2.30.10^{-4}$

	ρ_0	$2(1-\rho_0)/N$	$s^2(a)$
$N = 100$	0.9	$2.00.10^{-3}$	$2.94.10^{-3}$
	0.99	$2.00.10^{-4}$	$1.75.10^{-4}$
	0.999	$2.00.10^{-5}$	$1.96.10^{-5}$

	ρ_0	$2(1-\rho_0)/N$	$s^2(a)$
$N = 1000$	0.9	$2.00.10^{-4}$	$2.80.10^{-4}$
	0.99	$2.00.10^{-5}$	$2.05.10^{-5}$
	0.999	$2.00.10^{-6}$	$2.17.10^{-6}$

	ρ_0	$2(1-\rho_0)/N$	$s^2(a)$
$N = 10000$	0.9	$2.00.10^{-5}$	$2.44.10^{-5}$
	0.99	$2.00.10^{-6}$	$2.21.10^{-6}$
	0.999	$2.00.10^{-7}$	$2.09.10^{-7}$

Not surprisingly, the agreement is best for ρ_0 close to 1.

C Discrete Sampling

We are here considering the error variance in the case where we include the limit $\Delta t \gg \mathscr{T}$.

We replace (61) and (62) by

$$\overline{x} = \frac{1}{N} \sum_{\ell=0}^{N-1} x(\ell \Delta t) \tag{C1}$$

and

$$
\begin{aligned}
\sigma^2\{x\} &= \left\langle (\overline{x} - \langle x \rangle)^2 \right\rangle \\
&= \frac{1}{N} \sum_{\ell'=0}^{N-1} \frac{1}{N} \sum_{\ell''=0}^{N-1} \left\langle \left(x(\ell'\Delta t) - \langle x \rangle \right) \left(x(\ell''\Delta t) - \langle x \rangle \right) \right\rangle \\
&= \langle x^2 \rangle \frac{1}{N} \sum_{\ell'=0}^{N-1} \frac{1}{N} \sum_{\ell''=0}^{N-1} \rho\left((\ell'' - \ell')\Delta t \right),
\end{aligned}
\tag{C2}
$$

where we have used (60) for the last step.

By comparing (C2) and (62), we see that the last is a poor approximation to the first when $\Delta t \gg \mathscr{T}$ since the smallest, non-zero increment in the summation (C2) is then so large that the corresponding change in $\rho(\tau)$ cannot be considered small.

Applying the Fourier transform

$$\rho(\tau) = \int_{-\infty}^{\infty} S(f) e^{q\pi i f \tau} df \tag{C3}$$

to (C2), we get

$$
\begin{aligned}
\sigma^2\{x\} &= \langle x^2 \rangle \frac{1}{N} \sum_{\ell'=0}^{N-1} \frac{1}{N} \sum_{\ell''=0}^{N-1} \int_{-\infty}^{\infty} S(f) \exp\left(2\pi i f \Delta t (\ell'' - \ell')\right) df \\
&= \langle x^2 \rangle \int_{-\infty}^{\infty} S(f) df \left| \frac{1}{N} \sum_{\ell=0}^{N-1} e^{2\pi i f \Delta t \ell} \right|^2 \\
&= \langle x^2 \rangle \int_{-\infty}^{\infty} H_N(f) S(f) df,
\end{aligned}
\tag{C4}
$$

where

$$
H_N(f) = \frac{\sin^2(\pi f N \Delta t)}{N^2 \sin^2(\pi f \Delta t)}.
\tag{C5}
$$

We are concerned with situations where $N = T/\Delta t \gg 1$ in which limit $H_N(f)$ becomes proportional to a so-called delta-comb, which is a sum of equidistantly spaced delta functions:

$$
H_N(f) \approx \lim_{N \to \infty} H_N(f) = \frac{1}{T} \sum_{m=-\infty}^{\infty} \delta\left(f - \frac{m}{\Delta t}\right).
\tag{C6}
$$

We thus obtain the general expression

$$
\sigma^2\{x\} = \frac{\langle x^2 \rangle}{T} \sum_{m=-\infty}^{\infty} \int_{-\infty}^{\infty} S(f) \delta\left(f - \frac{m}{\Delta t}\right) df = \frac{\langle x^2 \rangle}{T} \sum_{m=-\infty}^{\infty} S\left(\frac{m}{\Delta t}\right),
\tag{C7}
$$

valid when $N \gg 1$.

In the limit $\Delta t \to 0$ there is only one term in the sum (C7) and we get

$$
\sigma^2\{x\} = \langle x^2 \rangle \frac{S(0)}{T},
\tag{C8}
$$

which is identical to (69).

In the other limit, $\Delta t \to \infty$, the sum becomes an integral:

$$
\begin{aligned}
\sigma^2\{x\} &= \frac{\langle x^2 \rangle}{T} \sum_{m=-\infty}^{\infty} S\left(\frac{m}{\Delta t}\right) \underbrace{\delta m}_{\equiv 1} \\
&= \langle x^2 \rangle \frac{\Delta t}{T} \sum_{m=-\infty}^{\infty} S\left(\frac{m}{\Delta t}\right) \delta\left(\frac{m}{\Delta t}\right) \\
&\approx \langle x^2 \rangle \frac{\Delta t}{T} \underbrace{\int_{-\infty}^{\infty} S(f) df}_{=1} = \frac{\langle x^2 \rangle}{N}.
\end{aligned}
\tag{C9}
$$

We see that now N_{eff} becomes equal to N as expected.

D Intermittent Sampling

In order to analyze intermittent time series, we first consider an uninterrupted time series $x_0(t)$ from which we can obtain an intermittent time series by

$$x(t) = \chi(t)x_0(t),$$

(D1)

where the two-valued *index function* $\chi(t)$ can attain the values one and zero.

Let us consider a more general system which can only exist in two states, an upper state and a lower. The probability that the system flips from one state to the other at least once in a given, small time interval is proportional to the duration Δt of this interval. Let this probability be $k\Delta t$, where k is a constant. Consequently, the probability that the system has not changed becomes $1 - k\Delta t$. The probability that the system has not changed state in the fixed, finite time $T = N\Delta t$ becomes

$$p(T) = \lim_{N\to\infty}\left\{(1 - k\Delta t)^N\right\} = \lim_{N\to\infty}\left\{\left(1 - k\frac{T}{N}\right)^N\right\} = e^{-kT}.$$

(D2)

The system we want to discuss has in general two different decay constants k_+ and k_-, characterizing flips from the upper to the lower and from the lower to the upper state, respectively.

We consider a system at start time $t = 0$ and seek the probability that the system is in the same state as the original one at time $t \geq 0$, in other words that it has flipped an even number of times. Let us start with the system in the upper state. Then the probability for no flips is

$$P_0^{++}(t) = e^{-k_+ t}.$$

(D3)

The probability for the system to be in the lower state at time t after one flip is

$$P_1^{+-}(t) = \int_0^t e^{-k_+ t_1} k_+ dt_1 e^{-k_-(t-t_1)}.$$

(D4)

Continuing this process, the probability for being back in the upper state after just two flips becomes

$$P_2^{++}(t) = \int_0^t e^{-k_+ t_1} k_+ dt_1 \int_{t_1}^t e^{-k_-(t_2-t_1)} k_- dt_2 P_0^{++}(t - t_2).$$

(D5)

We thus obtain the following recursion relation for $n > 1$

$$
\begin{aligned}
P_{2n}^{++}(t) &= \int_0^t e^{-k_+ t_1} k_+ dt_1 \int_{t_1}^t e^{-k_-(t_2-t_1)} k_- dt_2 P_{2n-2}^{++}(t - t_2) \\[2mm]
&= k_+ k_- \int_0^t e^{-(k_+ - k_-)t_1} dt_1 \int_{t_1}^t e^{-k_- t_2} P_{2n-2}^{++}(t - t_2) dt_2 \\[2mm]
&= \frac{k_+ k_-}{k_+ - k_-} \int_0^t \left\{ e^{-k_- t_1} - e^{-k_+ t_1} \right\} P_{2n-2}^{++}(t - t_1) dt_1,
\end{aligned}
$$

(D6)

where we have applied integration by parts to obtain the last expression.

We may now sum up all these probabilities:

$$P^{++}(t) \equiv \sum_{n=0}^{\infty} P_{2n}^{++}(t) = e^{-k_+ t} + \sum_{n=1}^{\infty} P_{2n}^{++}(t). \tag{D7}$$

Inserting (D6) we get the following integral equation for $P^{++}(t)$:

$$P^{++}(t) = e^{-k_+ t} + \frac{k_+ k_-}{k_+ - k_-} \int_0^t \left\{ e^{-k_- t_1} - e^{-k_+ t_1} \right\} P^{++}(t - t_1) \, dt_1. \tag{D8}$$

This is a Volterra integral equation of the second kind and it can easily be solved by means of Laplace transforms. Writing $\mathscr{P}^{++}(s) \equiv \mathscr{L}\left\{P^{++}(t)\right\}$ we obtain

$$\mathscr{P}^{++}(s) = \frac{1}{s + k_+} + \frac{k_+ k_-}{k_+ - k_-} \mathscr{P}^{++}(s) \left\{ \frac{1}{s + k_-} - \frac{1}{s + k_+} \right\} \tag{D9}$$

with the solution

$$\mathscr{P}^{++}(s) = \frac{1}{k_+ + k_-} \left\{ \frac{k_-}{s} + \frac{k_+}{s + k_+ + k_-} \right\}. \tag{D10}$$

Transforming back to time domain yields

$$P^{++}(t) = \frac{k_- + k_+ e^{-(k_+ + k_-)t}}{k_+ + k_-}. \tag{D11}$$

This result could have been obtained in a simpler way, as pointed out by Jakob Mann (2000, private communication), namely as the solution to the first-order differential equation

$$\frac{dP^{++}}{dt} = -k_+ P^{++}(t) + k_- \left(1 - P^{++}(t)\right). \tag{D12}$$

However, we have adopted the method described here because it has provided a useful insight in the 'mechanism of flips'.

The probability that the system is in the lower state when it is at the upper at $t = 0$ is of course

$$P^{+-}(t) = 1 - P^{++}(t) = \frac{k_+ - k_+ e^{-(k_+ + k_-)t}}{k_+ + k_-}. \tag{D13}$$

Similarly, we can start in the lower state and the corresponding probabilities $P^{--}(t)$ and $P^{-+}(t)$ are obtained by interchanging k_+ and k_- in (D11) and (D13).

We may now proceed to determine the expected number of double flips in the time t. A priori we would expect this number to be proportional to t, but we need a rigorous proof that this is indeed the case. Let us again consider the system in the upper position at time $t = 0$. Then the number $N^+(t)$ of double flips can be expressed in terms of $P^{++}(t)$ as follows:

$$N^+(t) = \sum_{n=0}^{\infty} n P_{2n}^{++}(t) = P_2^{++}(t) + \sum_{n=2}^{\infty} n P_{2n}^{++}(t). \tag{D14}$$

Inserting (D6) we get

$$
\begin{aligned}
N^+(t) \;=\; & P_2^{++}(t)+ \\
& \frac{k_+k_-}{k_+-k_-}\int_0^t \left\{ \mathrm{e}^{-k_-(t-\theta)}-\mathrm{e}^{-k_+(t-\theta)} \right\} \mathrm{d}\theta \sum_{n=2}^{\infty} n P_{2n-2}^{++}(\theta).
\end{aligned}
\tag{D15}
$$

The sum in the integral can be rearranged as follows

$$
\begin{aligned}
\sum_{n=2}^{\infty} n P_{2n-2}^{++}(\theta) \;=\; & \sum_{n=2}^{\infty} (n-1)P_{2n-2}^{++}(\theta)+\sum_{n=2}^{\infty} P_{2n-2}^{++}(\theta) \\
\;=\; & N^+(\theta)+P^{++}(\theta)-P_0^{++}(\theta)
\end{aligned}
\tag{D16}
$$

and, applying this expression, (D15) becomes

$$
N^+(t) = \frac{k_+k_-}{k_+-k_-}\int_0^t \left\{ \mathrm{e}^{-k_-(t-\theta)}-\mathrm{e}^{-k_+(t-\theta)} \right\}\left\{N^+(\theta)+P^{++}(\theta)\right\}\mathrm{d}\theta,
\tag{D17}
$$

where we have used (D6) for $n=1$. Again we take the Laplace transform, use (D10), and solve for $\mathcal{N}^+(s) \equiv \mathscr{L}\left\{N^+(t)\right\}$. Thus

$$
\begin{aligned}
\mathcal{N}^+(s) \;=\; & \frac{k_+k_-(k_+-k_-)}{(k_++k_-)^3}\left\{\frac{1}{s}-\frac{1}{s+k_++k_-}\right\} \\
& +\frac{k_+k_-}{(k_++k_-)^2}\left\{\frac{k_-}{s^2}-\frac{k_+}{(s+k_++k_-)^2}\right\}
\end{aligned}
\tag{D18}
$$

with the inverse transform

$$
\begin{aligned}
N^+(t) \;=\; & \frac{k_+k_-(k_+-k_-)}{(k_++k_-)^3}\left\{1-\mathrm{e}^{-(k_++k_-)t}\right\} \\
& +\frac{k_+k_-}{(k_++k_-)^2}\left\{k_-t-k_+t\mathrm{e}^{-(k_++k_-)t}\right\}
\end{aligned}
\tag{D19}
$$

Similarly, the average number of double flips if the system starts in the lower state becomes

$$
\begin{aligned}
N^-(t) \;=\; & \frac{k_+k_-(k_--k_+)}{(k_++k_-)^3}\left\{1-\mathrm{e}^{-(k_++k_-)t}\right\} \\
& +\frac{k_+k_-}{(k_++k_-)^2}\left\{k_+t-k_-t\mathrm{e}^{-(k_++k_-)t}\right\}.
\end{aligned}
\tag{D20}
$$

The total expected number $N(t)$ of double flips, irrespective of the initial state can be determined from (D19) and (D20) by adding them, after weighting them with the a priori

probabilities Π^+ and Π^+ for being in the upper state and the lower state, respectively. Since $\Pi^+ k+ = \Pi^- k-$ they become

$$\Pi^+ = \frac{k_-}{k_+ + k_-} \tag{D21}$$

and

$$\Pi^- = \frac{k_+}{k_+ + k_-}. \tag{D22}$$

The result is

$$
\begin{aligned}
N(t) &= \frac{k_-}{k_+ + k_-} N^+(t) + \frac{k_+}{k_+ + k_-} N^-(t) \\
&= \frac{k_+ k_-}{(k_+ + k_-)^3} \left\{ \left(k_+^2 + k_-^2\right) t - 2 k_+ k_- t \mathrm{e}^{-(k_+ + k_-)t} \right\} \\
&\quad - \frac{k_+ k_- (k_+ - k_-)^2}{(k_+ + k_-)^4} \left\{ 1 - \mathrm{e}^{-(k_+ + k_-)t} \right\}.
\end{aligned} \tag{D23}
$$

We see that only for large values of t can the average number of double flips be considered proportional to t:

$$N(t) \approx \frac{k_+ k_- \left(k_+^2 + k_-^2\right) t}{(k_+ + k_-)^3}, \qquad (k_+ + k_-)t \gg 1. \tag{D24}$$

When t is small we get

$$N(t) \approx \frac{1}{2} k_+ k_- t^2, \qquad (k_+ + k_-)t \ll 1. \tag{D25}$$

We now assign the value 1 to the upper state and 0 to the lower. In order to relate the decay constants k_+ and k_- to observable quantities we calculate the ensemble mean and ensemble variance of the corresponding, stationary time series $\chi(t)$. We have

$$\langle \chi(t) \rangle = \Pi^+ \times 1 + \Pi^- \times 0 = \frac{k_-}{k_+ + k_-} \tag{D26}$$

and

$$\sigma_\chi^2 \equiv \langle (\chi(t) - \langle \chi(t) \rangle)^2 \rangle == \Pi^+ \times 1 - \langle \chi(t) \rangle)^2 \rangle = \frac{k_+ k_-}{(k_+ + k_-)^2}. \tag{D27}$$

We may use observation of $\chi(t)$ over a long time to determine k_+ and k_- in terms of σ_χ^2 and the long-term rate of double flips

$$\eta = \frac{k_+ k_- \left(k_+^2 + k_-^2\right)}{(k_+ + k_-)^3}. \tag{D28}$$

Solving the two equations (D27) and (D28) with respect to k_+ and k_-, we get, if we assume that $k_- \geq k_+$:

$$k_\pm = \frac{\eta}{2\sigma_\chi^2 \left(1 - 2\sigma_\chi^2\right)} \left\{ 1 \mp \sqrt{1 - 4\sigma_\chi^2} \right\}. \tag{D29}$$

The auto-covariance function for $\chi(t)$ is

$$
\begin{aligned}
R_\chi(t_2 - t_1) &\equiv \left\langle \left(\chi(t_1) - \langle\chi\rangle\right)\left(\chi(t_2) - \langle\chi\rangle\right)\right\rangle \\
&= \left\langle \chi(t_1)\chi(t_2)\right\rangle - \langle\chi\rangle^2 = \Pi^+ P^{++}(t_2 - t_1) - \langle\chi\rangle^2 \\
&= \frac{k_+ k_-}{(k_+ + k_-)^2}\, \mathrm{e}^{-(k_+ + k_-)|t_2 - t_1|}.
\end{aligned}
\tag{D30}
$$

We now define the total time Θ during which $\chi(t)$ is equal to one during the observation time T by

$$
\Theta(T) = \int_0^T \chi(t)\,\mathrm{d}t.
\tag{D31}
$$

This definition immediately implies that

$$
\langle\Theta(T)\rangle = \frac{k_- T}{k_+ + k_-}.
\tag{D32}
$$

The variance becomes

$$
\begin{aligned}
\sigma_\Theta^2(T) &\equiv \left\langle \left(\Theta(T) - \langle\Theta(T)\rangle\right)^2 \right\rangle \\
&= \left\langle \int_0^T \left(\chi(t_1) - \langle\chi\rangle\right)\mathrm{d}t_1 \int_0^T \left(\chi(t_2) - \langle\chi\rangle\right)\mathrm{d}t_2 \right\rangle \\
&= \int_0^T \mathrm{d}t_1 \int_0^T \mathrm{d}t_2\, R_\chi(t_2 - t_1) \\
&= \frac{k_+ k_-}{(k_+ + k_-)^2}\int_0^T \mathrm{d}t_1 \int_0^T \mathrm{d}t_2\, \mathrm{e}^{-(k_+ + k_-)|t_2 - t_1|} \\
&= \frac{2 k_+ k_- T}{(k_+ + k_-)^3}\left\{ 1 - \frac{1 - \mathrm{e}^{-(k_+ + k_-)T}}{(k_+ + k_-)T} \right\}.
\end{aligned}
\tag{D33}
$$

Since

$$
\frac{\sigma_\Theta(T)}{\langle\Theta(T)\rangle} \approx \sqrt{\frac{2 k_+/k_-}{(k_+ + k_-)T}}
\tag{D34}
$$

when $(k_+ + k_-)T \gg 1$, we find that $\Theta(T)$, given by (D31), in this limit is a good approximation to the ensemble average $\langle\Theta(T)\rangle$.

The sample mean of $x(t) = \chi(t)x_0(t)$ as given in (D1 now becomes

$$
\bar{x} = \frac{1}{\Theta}\int_0^T x(t)\,\mathrm{d}t = \frac{1}{\Theta}\int_0^T \chi(t)x_0(t)\,\mathrm{d}t.
\tag{D35}
$$

Although this is certainly not true in general we assume—in order to capture in a simple way the essence of the consequences of conditional sampling—$\chi(t)$ and $x_0(t)$ uncorrelated.

Consequently, the ensemble mean of $\overline{x}$ becomes

$$
\begin{aligned}
\langle \overline{x} \rangle &= \left\langle \frac{1}{\Theta} \int_0^T \chi(t) x_0(t) \, dt \right\rangle \\
&= \int_0^T \left\langle \frac{\chi(t)}{\Theta} \right\rangle \langle x_0(t) \rangle \, dt = \langle x_0 \rangle = 0.
\end{aligned}
\tag{D36}
$$

For the variance of $\overline{x}$ we get

$$
\begin{aligned}
\sigma^2\{x\} &= \langle (\overline{x} - \langle \overline{x} \rangle)^2 \rangle = \left\langle \frac{1}{\Theta} \int_0^T \chi(t_1) x_0(t_1) \, dt_1 \frac{1}{\Theta} \int_0^T \chi(t_2) x_0(t_2) \, dt_2 \right\rangle \\
&\approx \frac{1}{\Theta} \int_0^T dt_1 \frac{1}{\Theta} \int_0^T dt_2 \langle x_0^2 \rangle \rho_0(t_2 - t_1) \langle \chi(t_1) \chi(t_2) \rangle \\
&= \frac{\langle x_0^2 \rangle}{\Theta} \int_0^T dt_1 \frac{1}{\Theta} \int_0^T dt_2 \, \rho_0(t_2 - t_1) \left\{ \langle \chi \rangle^2 + R_\chi(t_2 - t_1) \right\},
\end{aligned}
\tag{D37}
$$

where $\rho_0(t)$ is the autocorrelation function for $x_0(t)$.

Combining this equation with (D26), (D27), (D30), and (D32), we see that

$$
\sigma^2\{x\} = \frac{\langle x_0^2 \rangle}{T} \int_0^T dt_1 \frac{1}{T} \int_0^T dt_2 \, \rho_0(t_2 - t_1) \left\{ 1 + \frac{k_+}{k_-} e^{-(k_+ + k_-)|t_2 - t_1|} \right\}.
\tag{D38}
$$

Assuming

$$
\rho_0(t) = e^{-|t|/\mathscr{T}},
\tag{D39}
$$

the error variance becomes in the limit where $(k_+ + k_-)T \gg 1$ as well as $T/\mathscr{T} \gg 1$

$$
\begin{aligned}
\sigma^2\{x\} &= 2\langle x_0^2 \rangle \frac{\mathscr{T}}{T} \left\{ 1 + \frac{k_+/k_-}{1 + (k_+ + k_-)\mathscr{T}} \right\} \\
&= 2\langle x_0^2 \rangle \frac{\mathscr{T}}{T} \left\{ 1 + \frac{\sigma_\chi^2/\langle \chi \rangle^2}{1 + \dfrac{\eta \mathscr{T}}{\sigma_\chi^2 (1 - 2\sigma_\chi^2)}} \right\},
\end{aligned}
\tag{D40}
$$

where the parameters in the last expression, defined by (D26), (D27) and (D28), can be derived directly from the measured record $\chi(t)$.

Acknowledgements

References

Box, G. E. P. & Muller, M. E. (1958), 'A note on the generation of random normal deviates', *Ann. Math. Statist.* **29**, 610–611.

Brazier, C. E. (1914), Recherches expérimentales sur les moulenets anémometrique, *in* 'Ann. Bur. Centr. Météorol. France', pp. 157–300.

Kanji, G. K. (1999), *100 Statistical Tests*, Sage Publications, London, Thousand Oaks, and New Delhi.

Kristensen, L. (1994), Cups, props and vanes, Technical Report R–766(EN), Risø National Laboratory.

Kristensen, L. (1998), 'Cup anemometer behavior in turbulent environments', *J. Atmos. Ocean. Technol.* **15**, 5–17.

Kristensen, L. (1999), 'The perennial cup anemometer', *Wind Energy* **2**, 59–75.

Kristensen, L. (2000), 'Measuring higher-order moments with a cup anemometer', *J. Atmos. Ocean. Technol.* **17**, 1139–1148.

Kristensen, L., Rathmann, O. & Hansen, S. O. (1999), Extreme winds in Denmark, Technical Report R–1068(EN), Risø National Laboratory.

Lenschow, D. H., Mann, J. & Kristensen, L. (1994), 'How long is long enough when measuring fluxes and other turbulence statistics', *J. Atmos. Ocean. Technol.* **11**, 661–673.

Middleton, W. E. K. (1969), *Invention of Meteorological Instruments*, The Johns Hopkins Press, Baltimore, MD.

Patterson, J. (1926), 'The cup anemometer', *Trans. Roy. Soc. Canada, Ser. III* **20**, 1–54.

Wyngaard, J. C. (1981), 'Cup, propeller, vane, and sonic anemometers in turbulence research', *Ann. Rev. Fluid Mech.* **13**, 399–423.

Bibliographic Data Sheet   **Risø–R–1218(EN)**

Title and author(s)

Field Calibration of Cup Anemometers

Leif Kristensen, Gunnar Jensen, Arent Hansen, and Peter Kirkegaard

ISBN	ISSN
87–550–2772–5; 87–550–2773–3 (Internet)	0106–2840

Dept. or group	Date
Department of Wind Energy	January 23, 2001

Groups own reg. number(s)	Project/contract No.
1105016–00	1105-16-00

Sponsorship

Pages	Tables	Illustrations	References
42	3	12	12

Abstract (Max. 2000 char.)

An outdoor calibration facility for cup anemometers, where the signals from 10 anemometers of which at least one is a reference can be can be recorded simultaneously, has been established. The results are discussed with special emphasis on the statistical significance of the calibration expressions. It is concluded that the method has the advantage that many anemometers can be calibrated accurately with a minimum of work and cost. The obvious disadvantage is that the calibration of a set of anemometers may take more than one month in order to have wind speeds covering a sufficiently large magnitude range in a wind direction sector where we can be sure that the instruments are exposed to identical, simultaneous wind flows. Another main conclusion is that statistical uncertainty must be carefully evaluated since the individual 10 minute wind-speed averages are not statistically independent.

Descriptors INIS/EDB

WIND; VELOCITY; ANEMOMETER; CALIBRATION; FIELD TESTS; DATA ANALYSIS