Final author response

Journal: AMT
Title: Field intercomparison of prevailing sonic anemometers
Author(s): Matthias Mauder and Matthias J. Zeeman MS No.: amt-2017-284 MS Type: Research article

We are grateful for the very valuable and constructive comments by the two reviewers. The comments are in black font, responses are presented in blue font and changes in the manuscript in red font.

Response to RC1 (by John Frank)

John Frank, USDA Forest Srvice In this manuscript seven sonic anemometers from six non-orthogonal designs and four manufactures are compared during a two and a half-week study at the TERENO/ICOS site in southern Germany. Half-hourly mean wind velocity, mean and standard deviation of temperature, standard deviation of vertical wind velocity, friction velocity, and buoyancy flux were compared between the seven anemometers. In general, all anemometers were reasonably similar, with the largest discrepancy being temperature measurements with the Gill anemometers.

The topic of this study is timely, with a growing interest in the accuracy in sonic anemometer measurements. The work presented here is clear, convincing, and thorough. I believe there is one main comment and a few minor issues that need to be addressed before it is acceptable for publication in Atmospheric Measurement Techniques.

I have one main comment that should be addressed. The discussion states "it is also possible that the flow distortion errors were very small for our experimental set-up because the angles of attack are close to being surface-parallel" with support from Figure 9 showing that most angles were within ±6° with a majority falling within an even narrower window. I am concerned that there is a high likelihood that each of these non-orthogonal anemometers have erroneous and unpredictable measurements of σw for winds with very small angles-of-attack. This is described in Appendix 2 of Frank et al. (2016a) where it is demonstrated for the CSAT3 that as the angle of-attack approaches 0° (i.e., near surface-parallel) that systematic w measurement uncertainties approach ±∞. This finding could be extended to any of the non-orthogonal anemometers included in this study, and I have included an Appendix at the end in this review to expand upon this topic. The discussion also states "The other interpretation that all anemometers are afflicted with the same bias appears less likely, since it is difficult to imagine that several instruments measure the same quantity equally wrong, despite the obvious differences in sensor geometry and internal data processing."

The figure and the text passages that this reviewer refers to are not part of the current manuscript published in AMTD. Based on the reviewer's comments in the Quick Review, we have double-checked our results and we found that there was a mistake in the calculation of the angles of attack. The spread in angles of attack during our experiment was actually much larger as previously thought, i.e. a standard deviation of 15° rather than ±6°. This standard deviation of 15° is at the upper end of values reported for previous intercomparison experiments. Therefore, we have also changed the line of arguments in the discussion in accordance with the new results. We now really believe that the different instruments all show the same biases despite their differences in geometry and internal corrections. Since there is strong evidence for a bias in σ_w of the CSAT3 from other studies (Horst et al. 2015, Frank et al. 2016, Huq et al. 2017), this seems to be the only logical conclusion. Nevertheless, we highly appreciate the reviewer's extensive comments on the problem of non-orthogonal anemometers, which he provided in the Appendix of RC1.

I suggest that "the other interpretation" might actually be true, that these systematic w measurement uncertainties that approach $\pm\infty$ could explain why these "instruments measure the same quantity equally wrong".

We now fully agree that these instruments measure the same quantity equally wrong. We have informed the editor about the changes that we have made after the Quick Review stage, but unfortunately, those notes did not reach the reviewer in time.

A key component of this experiment is the inclusion of the CSAT3, which does not apply any transducer shadowing correction, which has been modeled (Huq et al. 2017) and observed (Horst et al. 2015) to have transducer shadowing errors that when transformed into orthogonal coordinates lead to unpredictable measurements for near surfaceparallel winds (Frank et al. 2016a). It is nearly impossible to know exactly what happens for near surface-parallel winds, but a simple evaluation of the Kaimal correction (Kaimal et al. 1990) applied to the CSAT3 yields a range of ± ~4° where systematic measurement errors approach $\pm\infty$ as shown in Figure 3f of Frank et al. (2016a). Because most angles were within ±6° in this study, I consider it extremely likely that the CSAT3 data has such errors. At the same time, all other anemometers in this study are probably more susceptible to this problem because their transducers are all tilted closer to the horizontal plane (45° for the Gill, R. M. Young and Metek versus 60° for the CSAT3).

I have two suggestions that can help address this issue. First, the authors could do a sensitivity analysis to quantify the potential impact of transducer shadowing on the CSAT3 σw and buoyancy flux measurements. I would suggest using both the piecewise (Kaimal et al. 1990) and sinusoidal (Wyngaard and Zhang 1985) corrections presented in the following appendix. While it is important to note that there is no consensus that these corrections are accurate (two studies have shown that they might account for about half of the shadowing (Frank et al. 2016b, Huq et al. 2017)), this will help to evaluate the statement that "the flow distortion errors were very small".

Again, this is a quotation from an earlier version of the manuscript, which is not part of the manuscript that is under review of this open discussion and published in AMTD. In contrast to the earlier version, we do not believe anymore that the flow distortion errors were very small, but rather that all tested anemometers are afflicted with a very similar flow distortion error. Therefore, we have even further strengthened the following statement accordingly:

A common significant systematic error of all tested instruments is quite possible, as suggested by Frank et al. (2016).

However, we find that this suggestion is extremely interesting for follow-up work, but we believe this is out of the scope of this manuscript. We acknowledge the need for an angle of attack dependent correction, but this should probably include data from multiple sites with different surface and vegetation properties.

Second, in conjunction with the histogram in Figure 9, an analysis of the relative contribution of winds from each angle of attack bin to the total σw and buoyancy flux measurements would be useful. Figure 9 currently shows that most data is in the bins -1° to 0° and 0° to 1°. How important are the measurements in these bins to the total σw and buoyancy flux measurements reported in figures 5 and 8? What is the contribution of winds that exist within -4° to 4° (i.e., a range over which the Kaimal correction as applied to the CSAT3 could conceivably result in unpredictable measurements)?

I very much look forward to the authors' critical evaluation of this topic as I believe it will be an immense benefit to the research community to thoroughly discuss non-orthogonal wind measurements.

We belief this is a misunderstanding, which stems from the fact the reviewer refers to an earlier version, which was altered during the Quick Review stage. Since the old Figure 9 was erroneous and is not included in the manuscript under discussion, we do not believe there is a need for further clarification here.

I have a minor comment about interpreting results based on offset/bias versus slope differences. My impression is that the authors focus more on offset/bias differences and less about slope. One example is on page 12, lines 4-5 "the fluctuations of sonic temperature agree much better". When I compared Tables 3 to 5, my attention immediately focused on the slopes, which are not much different. The average absolute difference from 1.00 (i.e., an extremely simple metric to summarize the group differences) was 4.0% for Table 3 (i.e., 1.05, 0.97, 1.01, 1.05, 1.06, 1.04 -> +5%, -3%, +1%, +5%, +6%, +4% -> (5+3+1+5+6+4)/6 = 4%) and 3.5% for Table 5. Similarly, on page 12, lines 12-14 it is commented the high slope of 1.06 "might be a direct consequence of the almost equally high regression slope of 1.05". This is not surprising, because ideally, slope errors in measuring Ts should commute to slope errors in σTs.

We agree that for flux measurements, an error in the slope is more severe that an error in the bias, since mean is always subtracted when calculating a covariance. The first sentence quoted in this comment refers to the Gill instruments. We think then this is a fair statement, because they have quite large deviations of the mean sonic temperatures, but its standard deviation appears comparable. We have modified this sentence for clarification:

Despite the large discrepancies of the mean sonic temperature measurements of the Gill instruments, the fluctuations of sonic temperature agree much better

We have also made the second quoted sentence stronger, because there is indeed a direct relation between the slope of the mean and the slope of the standard deviation:

The METEK.uSonic.omni stands out because it has the highest regression slope of 1.06, which is a direct consequence of the almost equally high regression slope of 1.05 for the mean sonic temperature measurement.

Specific comments:

The introduction is excellent, and one of the better that I've read for sonic anemometer studies. What is the sampling rate of the sonic anemometers?

This information has been added to the manuscript.

The sampling rate was 20 Hz, except for the CSAT3_2, which was sampled at 60 Hz, and the Gill_HS, which was sampled at 10 Hz.

Figure 2: It would be good to mention either on the figure or in the caption that the model of anemometer corresponds to the same names listed in Table 1.

We agree and we have added that information in the caption of Figure 2.

they are presented from left to right in the same order as they are listed in Table 1.

Page 7, line 15: Are the obstructed wind directions based on 30-minute mean direction or some other metric?

They are indeed based on 30-minute mean wind directions. We have added this information.

These quantities were filtered for rain (during the respective half hour or the half hour before as recorded by a Vaisala WXT520 sensor of the nearby TERENO station DEFen), obstructed wind directions φ based on 30-minute averages (70° < φ < 110°; 250° < φ < 290°) and non-steady-state conditions, …

2.2 Data processing: The software R should be cited, with details provided by the R function “citation()”.

This citation of the R software has been added.

Table 2 and Figure 3: There is a slight difference in nomenclature between “mean wind velocity” and “total wind velocity”. Based on the 2-D rotation, these should be the same, but consistent labels would be good.

We agree and we have now consistently used the term “mean total wind velocity”.

Table 2 and 4. The slope difference between the CSAT3 anemometers is actually quite large (1.02 versus 0.97 for U and 0.98 versus 1.03 for σw). Do the authors believe that this is due more to the repeatability of measurements using the same anemometer design or to the 54 m separation distance?

We cannot really answer this question based on our results. Based on the experience from other intercomparison experiments, differences are on the order of 2 - 3% are certainly possible between the different locations but can also be possible due to instrumental differences. Moreover, we would like to stress that the biases and RMSDs are really small despite the slightly larger slopes.

Tables 2-7 and Figures 3-8: I could imagine all information from these tables could be moved to the blank space of the corresponding figures, thus saving space in the manuscript. I would also mention that something like R-squared values might be very useful to include as well.

The final layout will be done be the publishing company and we will then check the galley-proofs to avoid blank space. We decided against using R-squared values because we find this metric sort of redundant when RMSD and bias values are already provided. These two metrics also provide more specific information as to what degree the differences are systematic or random, while R-squared lumps both differences together.

Tables 3, 5, and 6: What was the metric used to identify "unusually large deviations from the etalon"? This should be mentioned in the methodology.

We agree that these criteria should be clarified and we added this information in the respective table captions:

Table 3: (slopes deviating more than 5% from unity and absolute differences of more than 1 K)

Table 5: (slopes deviating more than 5% from unity and absolute differences larger than 0.05 K)

Table 6: (slopes deviating more than 5% from unity)

Page 12, Line 14-15: What does "agreement between the two CSAT3 is except for a few outliers" mean?

Thanks, two words were missing here:

The agreement between the two CSAT3 is very good except for a few outliers

Page 13, line 4: "Yount" should be "Young".

This has been corrected.

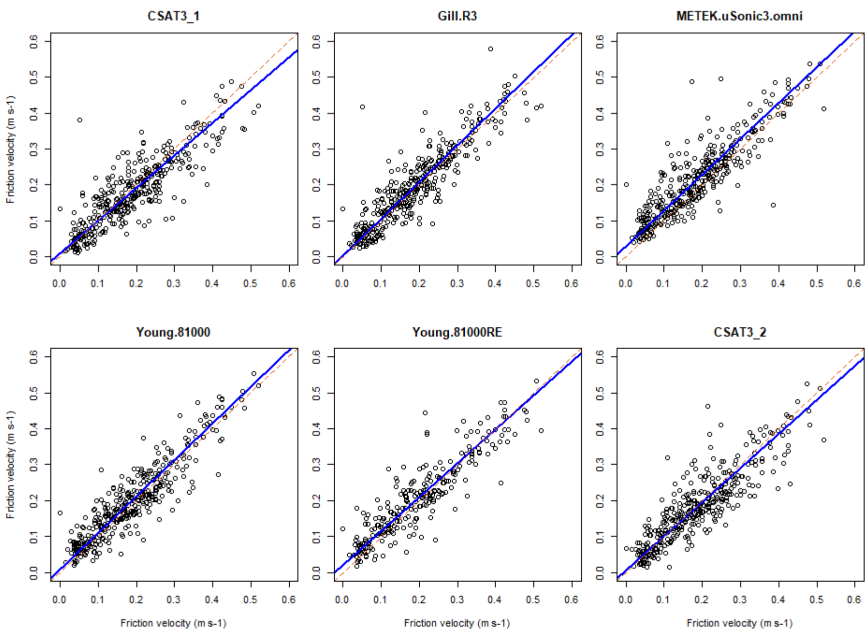
Page 13, line 11-12: It could be noted that CSAT3_2 is the second lowest, so both CSAT3 are fairly low.

We agree and we have added the following sentence:

Similarly, the CSAT_2 shows the second lowest regression slope, but its bias and RMSD is very similar to the other instruments.

Figure 7: One of the ylabels are missing.

Thanks, we have recompiled this figure including the ylabel:



Page 16, line 6: “error of due” should be “error due”.

Thanks, we have removed the word “of”

Figure 9: Are the magnitudes of the numbers of y-axis correct? The number of occurrences seems extremely low for instantaneous 10 Hz or 20 Hz data. Or is this half-hour average angle of-attack? The number of occurrences seems similar to the number of half-hours. If this is the case, then all text that refers to “small … angles of attack” (beginning with page 16, lines 25-26) must be revised to reflect the instantaneous angles experienced by the anemometers.

Indeed, the values were too low due to a mistake in the data analysis routine. We have therefore removed this figure already after the Quick Review because this whole line of argument saying that we had exceptionally low flow angles and therefore low systematic errors does not hold anymore. For completeness, we now provide the standard deviation of the flow angle, which is 15°, ranging at the upper end of values reported in the literature. Thanks to the reviewer’s remark in the Quick Review, we were already able to correct this mistake.

Page 18, line 11-14: The discussion of “type A” and “type B” should occur earlier in the paper than the conclusion.

We have added the following sentence to the discussion section:

Now, all tested instruments are within the limits that Mauder et al. (2006) classified as type A, i.e. sonic anemometers suitable for fundamental turbulence research.

Response to RC2 (anonymous)

General Comments "As the last comprehensive intercomparison experiments were conducted more than 10 years ago, …" is the motivation for the authors do carry out a new intercomparison for prevailing sonic anemometers. They present the analysis and the results in a well-prepared manuscript in a straight and standard way. The sonic-user community will be eager to see how the different instrument types perform. Insofar it is worthwhile to publish their results and it is perfectly within the scope of AMT. There are however some points the authors should address.

The explanation/discussion of the much better agreement is not convincing. I don't understand what the consistent digital data acquisition has to do with the better agreement. Give an example. And demonstrate how your quality tests improve the agreement.

The quality tests of Mauder et al. (2013) are the basis for the quality control applied for this intercomparison. The effect of these tests is presented in that paper as well, also for the same site. In addition, we applied a filter for rain and obstructed wind directions. Our wind sector filtering implies the exclusion of all obstructions, including the tripod and the neighboring systems at a distance of 9 m. In addition, the filtering for rainy periods was critical to exclude implausible measurements. In the discussion section of revised version, we have provided some examples of the comparison statistics for the dataset after only the tests of Mauder et al. (2013) had been applied and also for the dataset after exclusion of obstructed wind sectors before the filtering for rainy periods was applied.

On top of that, the filtering for obstructed wind direction sectors and for rain, as described in section 2.2, was crucial to remove poor quality data. Both additional steps improved the agreement between instruments considerably. For σ_w, regression slopes ranged between 1.00 and 1.24 and intercepts were between −0.05 and 0.00 m s^{-1} after processing according to Mauder et al. (2013). After filtering for obstructed wind direction, slopes ranged between 0.98 and 1.22 and intercepts remained between −0.05 and 0.00 m s^{-1}. As can be seen from the results (Table 4), the overall agreement further improved after the filtering for rainy periods. Especially, some outliers of the CSAT3_2, which did not have the rain-guard meshes at the transducer heads, were rejected after this step. The effect of the data filtering on other quantities, such as Hs, was smaller. Here, the slopes ranged already only between 0.97 and 1.00 after processing according to Mauder et al. (2013), which did not change much further after filtering for obstructed wind directions and for rainy periods (Table 7). This can be explained by the fact that the scheme of Mauder et al. (2013) is designed for quality control of fluxes and not necessarily standard deviations. It therefore much stricter for H_s than for σ_w.

You also indicate that contributions in changes of firmware might have an influence. Say more about that. And finally you say that five instruments apply "some sort" of correction. There is more information about the corrections, they should not be treated as blackbox. Add the information where it is existiong. E.g. the calibration files for R3 and HS are available and can be applied later (at least it used to be like that). One has the possibility to sample uncalibrated data and apply the calibration afterwards. Did you do that? The HS and the CSAT3: how would they compare then? The sonics from Young and Metek seem to be black boxes but they allow to switch on and off a wake or head correction. You probably used the sonics always with the corrections on. Any idea how strong the corrections are? It irritates me that an instrument like the Young 81000 with a magic wake correction is so close to the other instruments. Insofar I can understand that you find the good agreement "somewhat surprising" and I can follow your conclusion that this is rather a conservative estimate because of special conditions (small variation in angles-of-attac. I guess you investigated the differences on azimuthal dependencies).

It is possible to operate the Gill instruments in an uncalibrated mode, but our intention was to compare the anemometers in the configuration recommended by the manufacturer. We do not have information about a calibration of the Gill instruments that can be applied during post-processing. We also used the METEK with the head correction turned on. As far as we know, the CSAT3 is the only one of the tested instruments that does not apply any internal correction at the sensor level. We do not have all details about the corrections applied by the different manufacturers. Therefore, the scope of this comparison was limited to a characterization of a typical configuration as it is applied by most users. We have added text providing details about the settings and firmware versions.

All other settings were left at the factory-recommended values, including flow-distortion corrections. The differences due to different firmware versions are quite well documented for the CSAT3. Accordingly to Burns et al. (2012), discrepancies between firmware versions 3 and 4 occur mostly for the sonic temperature measurement and they become significant for wind speeds larger than 8 m s^{-1}. During our field campaign, wind speeds were mostly lower than 5 m s^{-1} (Figure 4). Therefore, we do not expect large errors. Nevertheless, we used the same firmware version (ver4) for both CSAT3.

We as authors share the reviewer's irritation about missing or incomplete information on the sensor-based corrections in the respective manuals or firmware documents by all manufacturers. Campbell Scientific certainly has the more transparent policies with respect to the internal processing routines.

The angle-of-attack figure disappeared. It was surprising to see that the deviations from horizontal were that small (mostly within ±6°). Now there is a standard deviation of 15°. What happened?

Based on the reviewer's comments in the Quick Review, we have double-checked our results and we found that there was a mistake in the calculation of the angles of attack. The spread in angles of attack during our experiment was actually much larger as previously thought, i.e. a standard deviation of 15° rather than ±6°. This standard deviation of 15° is at the upper end of values reported for previous intercomparison experiments. Therefore, we have also changed the line of arguments in the discussion in accordance with the new results. We now really believe that the different instruments all show the same biases despite their differences in geometry and internal corrections. Since there is strong evidence for a bias in σ_w of the CSAT3 from other studies (Horst et al. 2015, Frank et al. 2016, Huq et al. 2017), this seems to be the only logical conclusion.

I do not see the advantage of using the PCA load in the first place, for deciding on an etalon. Choosing rather one instrument for all comparison is much more stringent and makes it easier to compare the instruments.

We wanted to avoid any subjectivity in choosing the reference instrument. If we had chosen the CSAT3 as etalon, as was done for previous intercomparsions, the presentation of the results would have been biased, perhaps favoring the CSAT3. Moreover, such a priori decision would contradict one of our main conclusions, that they all are equally suitable for flux measurements. Therefore, we would like to stick with the PCA-based decision on the etalon, because it allows us to compare the all results with the "best" estimate of a certain quantity.

Comparison plots are a bit monotone and do not transmit much information. Plotting rather differences to reference than sonic value versus sonic value gives an immediate impression on statistics. For a direct connection to the scatter plots the regression results should be placed in the plots. Special features can be highlighted in the text.

Thanks for this suggestion. We agree that there are other ways to present the results of such an intercomparison. However, we followed the style of previous studies, which have also used similar scatter plots (Dyer et al., 1982; Fratini and Mauder, 2014; Mauder et al., 2007; Tsvang et al., 1985) in order to maintain comparability. The monotone nature of the plots in our study, in comparison to plots in previous studies with much more scatter, is an actual result.

Comment on the speed-related temperature of a CSAT3 (Firmware v3)? You don't mention whether you determined the zero offset of the two CSAT3 before the experiment. Did you? The serial numbers of the CSAT3s tell us that they are relatively old instruments. How long ago was their last calibration? Figure 5: why the CSAT3 deviate that much although they should be better comparable. Could it be related to zero offsets or old calibration?

The older of the two CSAT3 (SN 0771) had been sent for re-calibration to the manufacturer in 2014. This one has firmware version 4t. The other one (SN 1791) still has its original manufacturer calibration; it was purchased in 2009. It has firmware version 4. We did not determine the zero offset before the experiment.

Technical corrections

Abstract 2/16 (Wieser et al., 2001). Full stop

Thanks, full stop has been added.

3/33 indications of a

This has been corrected.

4/31 synchronized how? Please be more specific how this was done? Why it is that important if you compare just average quantities?

Here, we wanted to stress the importance of digital data acquisition with precise clocks in order to attribute the correct time stamp to each data line. Therefore, more information on these details are provided in the revised version:

Data from all instruments were digitally recorded on synchronized single-board computers (BeagleBone Black, BeagleBoard.org Foundation, Oakland Twp, MI, USA), equipped with temperature-compensated clocks (Chronodot, Macetech LLC, Vancouver, WA, USA), using an event-driven protocol for recording data lines, implemented in the Python programming language. The digital recording minimizes the influence of data cable properties on signal quality and minimizes the impact of loss of resolution by conversion between analog and digital signals outside the scope of the sensor. Issues stemming from cable properties usually have a more apparent effect on digital than on analog signal transmissions. In case of a signal deterioration by oxidation of contacts or loosening cable connections, digitally transmitted data lines will start to show up in a corrupted format, while loss of signal resolution in analog transmission may go unnoticed for some time. Therefore, the potential for added uncertainty to the observations recorded by analog data transmission can in part be avoided by digital communications.

5/5 DE-Fen ? 6/5 It looks shaky. Were there guy wires?

The tripods did not have guy wires. Nevertheless, they are more stable than they might look like on the photograph because the legs are partially hidden within the grass canopy. We also checked the spectra of the wind velocities and found no visible deviations from the typical inertial sub-range behavior, which might indicate vibrations of the masts at distinct frequencies.

7/2 All data were processed

This has been corrected.

7/15 DE-Fen

The missing dash has been added.

8/6 total wind velocity? You mean the magnitude of the 3d vector i.e. =(u2 +v2 +w2) 0.5 compared to the horizontal wind speed =(u2 +v2) 0.5, which is your mean wind speed?

Since we have applied double rotation, the mean vertical wind velocity is zero and the mean cross-wind velocity is also zero. For clarification, we have now consistently used the term “mean total wind velocity”.

12/5 was chosen as etalon for

This has been corrected.

12/15 CSAT3 is very good except for

This has been corrected.

14/5 Young.81000RE

This has been corrected.

15/5 The lower row is slightly too large so the y-axis is missing

Thanks, reviewer 1 has also noticed this, and we have recompiled the figure.

16/3 of this study for many

This has been corrected.

16/7 etalon for this quantity because (is a redundant, or omit “For this comparison”)

This has been corrected.

17/7 error of due

This has been corrected.

17/6 measurements systems (?)

This has been corrected.

20/2 243-251 instead of 363-372

This has been corrected

21/6 Frank and Massman listed twice

Thanks for the careful check of the reference list. This mistake has been corrected.

Field intercomparison of prevailing sonic anemometers

Matthias Mauder[1], Matthias J. Zeeman[1]

[1]Karlsruhe Institute of Technology, Institute of Meteorology and Climate Research, Atmospheric Environmental Research, Garmisch-Partenkirchen, Germany

Correspondence to: Matthias Mauder (matthias.mauder@kit.edu)

Abstract. Three-dimensional sonic anemometers are the core component of eddy-covariance systems, which are widely used for micrometeorological and ecological research. In order to characterize the measurement uncertainty of these instruments we present and analyse the results from a field intercomparison experiment of six commonly used sonic anemometer models from four major manufacturers. These models include Campbell CSAT3, Gill HS-50 and R3, METEK uSonic-3 Omni, R.M. Young 81000 and 81000RE. The experiment was conducted over a meadow at the TERENO/ICOS site De-Fen in southern Germany over a period of 16 days in June of 2016 as part of the ScaleX campaign. The measurement height was 3 m for all sensors, which were separated by 9 m from each other, each on its own tripod, in order to limit contamination of the turbulence measurements by adjacent structures as much as possible. Moreover, the high-frequency data from all instruments were treated with the same post-processing algorithm. In this study, we compare the results for various turbulence statistics, which include mean horizontal wind speed, standard deviations of vertical wind velocity and sonic temperature, friction velocity and the buoyancy flux. Quantitative measures of uncertainty, such as bias and comparability, are derived from these results. We find that biases are generally very small for all sensors and all computed variables, except for the sonic temperature measurements of the two Gill sonic anemometers (HS and R3), confirming a known transducer-temperature dependence of the sonic temperature measurement. The best overall agreement between the different instruments was found for the mean wind speed and the buoyancy flux.

1 Introduction

Although sonic anemometers have been used extensively for several decades in micrometeorological and ecological research, there is still some scientific debate about the measurement uncertainty of these instruments. This is due to the fact that an absolute reference for the measurement of turbulent wind fluctuations in the free atmosphere does not exist. Traditionally, two approaches have been applied to evaluate the performance of sonic anemometers, either by placing them in a wind-tunnel and testing them for different flow angles, or by putting different instruments next to each other in the field over a homogeneous surface, so that all of them can be expected to measure the same wind velocities and turbulence statistics. The first approach has the advantage that the true flow characteristics are well known; however, the characteristics of the flow deviate far from those in the turbulent atmospheric surface layer where sonic-anemometers are typically deployed. Reynolds numbers in a wind tunnel, for instance, are several orders of magnitude smaller than under natural conditions. In contrast, the second intercomparison approach has the disadvantage, that it lacks an uncontested reference; however, such field experiments allow the simultaneous evaluation of several instruments under real-world conditions. In other words, the first approach has a high internal validity while the second approach has a high external validity.

Wind-tunnel experiments have been an important milestone towards revealing and quantifying probe-induced flow distortion effects. One of the first wind-tunnel tests including a correction equation for flow distortion effects is reported by Kaimal (1979). Considering the results of another wind-tunnel study about a three-dimensional hot-wire anemometer, Högström (1982) stressed the importance of such test for all turbulence sensors, and wind-tunnel experiments soon became a standard method for optimizing and calibrating sonic anemometers. Subsequently, Zhang et al. (1986) developed a new sonic anemometer based on measurements of from the wind-tunnel, which inspired the design of the Campbell CSAT3. A further wind-tunnel calibration for the Gill Solent R2 sonic anemometer is presented by Grelle and Lindroth (1994).

However, researchers soon realized that the transferability of wind-tunnel experiments to field conditions is limited. A very interesting comparative wind-tunnel study about several sonic anemometers (Gill Solent, METEK USA-1, Kaijo Denki TR-61A, TR-61B, and TR-61C) is conducted by Wieser et al. (2001). They evaluate flow distortion correction algorithms provided by the respective manufacturers and come to the following conclusion „Because of the very low level of turbulence in the wind tunnel (no fences or trip devices have been used), the size and stability of vortices set up behind struts may be increased in comparison with field measurements.” (Wieser et al., 2001). Moreover, Högström and Smedman (2004) present a critical assessment of laminar wind-tunnel calibrations by using a hot-film instrument as reference during a field experiment over a flat and level coastal area with very low vegetation. Their results indicate that wind-tunnel based corrections might be overcorrecting, or at least do not improve the comparison with the reference measurement of turbulence statistics.

Despite these known limitations, more extensive wind-tunnel calibration studies were conducted, which led to the publication of the so-called angle-of-attack correction for Gill Solent R2 and R3 (van der Molen et al., 2004; Nakai et al., 2006). However, it is often overlooked that angle-of-attack dependent errors might partially be an artefact of wind-tunnel experiments, because in quasi-laminar wind-tunnel flows the angle-of-attack remains constant. In contrast, the flow distorting caused by the same geometrical structure is much smaller under turbulent conditions, when the three-dimensional wind vector and the corresponding flow angles fluctuate constantly (Huq et al., 2017).

In order to address concerns about the validity of these wind-tunnel based calibrations, the angle-of-attack based flow distortion concept was investigated in the field under natural turbulent conditions. Nakai and Shimoyama (2012) mounted several Gill WindMaster instruments at different angles next to each other above a short grass canopy, and Kochendorfer et al. (2012) conducted a very similar field experiment focussing on RM Young Model 81000 anemometers, while the Campbell CSAT3 was only briefly examined. It has to be noted that the results of these two studies were interpreted under the false assumption that the instantaneous wind vector remains unchanged between different instruments that are mounted more than 1 m apart, which contradicts the concept of a fluctuating turbulent flow with a certain decay of the spatial autocorrelation function (Kochendorfer et al., 2013; Mauder, 2013).

However, such side-by-side comparisons with different alignment of the same instrument can be quite instructive, as long as only turbulence statistics are analysed, which can indeed be considered to be similar across several metres over homogeneous surfaces. Although their study site is less than ideal for a field intercomparison (over a sloped forest canopy within the roughness sublayer), Frank et al. (2013) found that non-orthogonal positioned transducers can underestimate vertical wind velocity (w) and sensible heat flux (H), by comparing the output of two pairs of CSAT3 anemometers while one pair was rotated by 90°. This finding was substantiated in a follow-up study (Frank et al., 2016), which also covers a side-by-side comparison of two CSAT3 mounted at different alignment angles plus two sonic anemometers with an orthogonal transducer array and a CSAT3 with one vertical path. An elaborate statistical analysis leads them to the conclusion: “Though we do not know the exact functional form of the shadow correction, we determined that the magnitude of the correction is probably somewhere between the Kaimal and double-Kaimal correction.” (Frank et al., 2016), referring to the original work of Kaimal (1979).

In a parallel chain of events, International Turbulence Comparison Experiments (ITCE) were carried out at different places around the world since the early days of sonic anemometry used for micrometeorological field campaigns (Dyer et al., 1982; Miyake et al., 1971; Tsvang et al., 1973, 1985), mostly with the aim to investigate the comparability of different instrumental designs. Typically, relative differences were analysed based on those comparative datasets, which generally suffer from the lack of a “true” reference measurement or etalon, but those experiments have the advantage that many anemometer models can be tested at once under real-world conditions. Nevertheless, also absolute biases were sometimes detected, such as the flow distortion from supporting structures, which from the 1976 ITCE was deduced from a non-zero mean vertical wind speed, especially for geometries with a supporting rod directly underneath the measurement volume (Dyer, 1981)~~(Dyer, 1981)~~.

In those early ITCEs, mostly custom-made instruments were tested. However, since the beginning of the 1990s, a growing number of commercial sonic anemometer models became available from a number of manufacturers. Based on their field intercomparison experiments, Foken and Oncley (1995) classified all instruments commonly used at the time according to their expected errors into those that are suitable for fundamental turbulence research and those that are sufficient for general flux measurements. About one decade later, several then popular models were compared in a thorough and comprehensive study by Loescher et al. (2005). They tested eight different probes for the accuracy of their temperature measurement in a climate chamber; they investigated biases of the *w*-measurement in a low-speed wind tunnel, and investigated differences in the turbulence statistics measured in the field. At about the same time, also Mauder et al. (2007) conducted a field intercomparison of seven different sonic anemometers as part of the international energy balance closure experiment EBEX-2000 above a cotton field in California. Both studies more or less confirmed the classification of Foken and Oncley (1995) who concluded that only the directional probes without supporting structure directly underneath the measurement volume meet the highest requirements of turbulence research, while no significant deviations between those top-class instruments were detected.

The persisting lack of energy balance closure at many sites around the world (Stoy et al., 2013) and the emerging indications of a general flux underestimation of non-orthogonal sonic arrays (Frank et al., 2013) are the primary motivation of a special field experiment by Horst et al. (2015). They conducted an intercomparison at an almost ideal site, which was flat, even and with a homogeneous fetch. Two CSAT3 representing a typical non-orthogonal sensor, were compared against two different orthogonal probes manufactured by Applied Technologies Inc. and one custom-made CSAT3 with one vertical path. Under the assumption that the flow-distortion correction of Kaimal (1979) is correct, they state that the CSAT3 requires a correction of 3% to 5%. This is in quite good agreement with the conclusion of Frank et al. (2016), who suggest a correction of the magnitude between Kaimal and double Kaimal, and the numerical study of Huq et al. (2017), which found an underestimation of 3% to 7%. Thus, at least for the CSAT3, some consensus is emerging about the magnitude of the correction required under turbulent conditions in the field.

Although the results on measurement error are converging for the CSAT3 model, less is known about the comparability between different sonic anemometer models available today. As the last comprehensive intercomparison experiments were conducted more than 10 years ago, and some new models have emerged on the market since then and some others have received firmware upgrades, we believe it is time for another field intercomparison covering commonly used sonic anemometers. We deployed six different models from four different manufacturers next to each other over a short grass canopy. Furthermore, two CSAT3 were tested simultaneously in order to compare the influence of transducer rainguards. An orthogonal regression analysis is applied to the turbulence statistics obtained from the different instruments, and quantitative measures of uncertainty, such as bias and comparability (RMSE), are derived.

2 Materials and methods

2.1 Field experiment

This sonic anemometer inter-comparison experiment took place at the Fendt field site in Southern Germany (DE-Fen, 47.8329°N 11.0607°E, 595 m a.s.l.), which belongs to the German Terrestrial Environmental Observatories (TERENO) network. The measurement period was from 06 June to 22 June 2016, and the intercomparison was conducted as part of the multi-scale field campaign ScaleX (Wolf et al., 2017), where the sonic anemometers were subsequently deployed at different locations. The landscape surrounding the site comprises gentle hills that are partially covered by forest (Figure 1), and the land cover within the footprint consisted of grassland with a canopy height of 0.25 m (Zeeman et al., 2017). The aerodynamic roughness length was estimated to be 0.03 m. In this field experiment, we compared seven sonic anemometers from four different manufacturers. A detailed list of all participating instruments is provided in Table 1.

Since the dominant wind direction is North for this site on typical summer days due to a thermal circulation between the Alps and the Alpine foreland (Lugauer and Winkler, 2005), we set-up all instrumented towers in a row from East to West. The sensors were separated by 9 m from each other in order to avoid flow distortion between neighboring towers. The measurement height of all sonic anemometers was 3.0 m, and they were oriented towards West (270°) for all non-omnidirectional probes (Figure 2). Data from all instruments were digitally recorded on synchronized single-board computers (BeagleBone Black, BeagleBoard.org Foundation, Oakland Twp, MI, USA), equipped with temperature-compensated clocks (Chronodot, Macetech LLC, Vancouver, WA, USA), using an event-driven protocol for recording data lines, implemented in the Python programming language. The digital recording minimizes the influence of data cable properties on signal quality and minimizes the impact of loss of resolution by conversion between analog and digital signals outside the scope of the sensor. Issues stemming from cable properties usually have a more apparent effect on digital than on analog signal transmissions. In case of a signal deterioration by oxidation of contacts or loosening cable connections, digitally transmitted data lines will start to show up in a corrupted format, while loss of signal resolution in analog transmission may go unnoticed for some time. Therefore, the potential for added uncertainty to the observations recorded by analog data transmission can in part be avoided by digital communications. ~~using an event-driven communication protocol implemented in Python programming language.~~ The sampling rate was 20 Hz, except for the CSAT3_2, which was sampled at 60 Hz, and the Gill_HS, which was sampled at 10 Hz. All other settings were left at the factory-recommended values, including flow-distortion corrections. The differences due to different firmware versions are quite well documented for the CSAT3. Accordingly to Burns et al. (2012), discrepancies between firmware versions 3 and 4 occur mostly for the sonic temperature measurement and they become significant for wind speeds larger than 8 m s^{-1}. During our field campaign, wind speeds were mostly lower than 5 m s^{-1} (Figure 4). Therefore, we do not expect large errors. Nevertheless, we used the same firmware version (ver4) for both CSAT3.

Formatiert: Hochgestellt

Formatiert: Hochgestellt

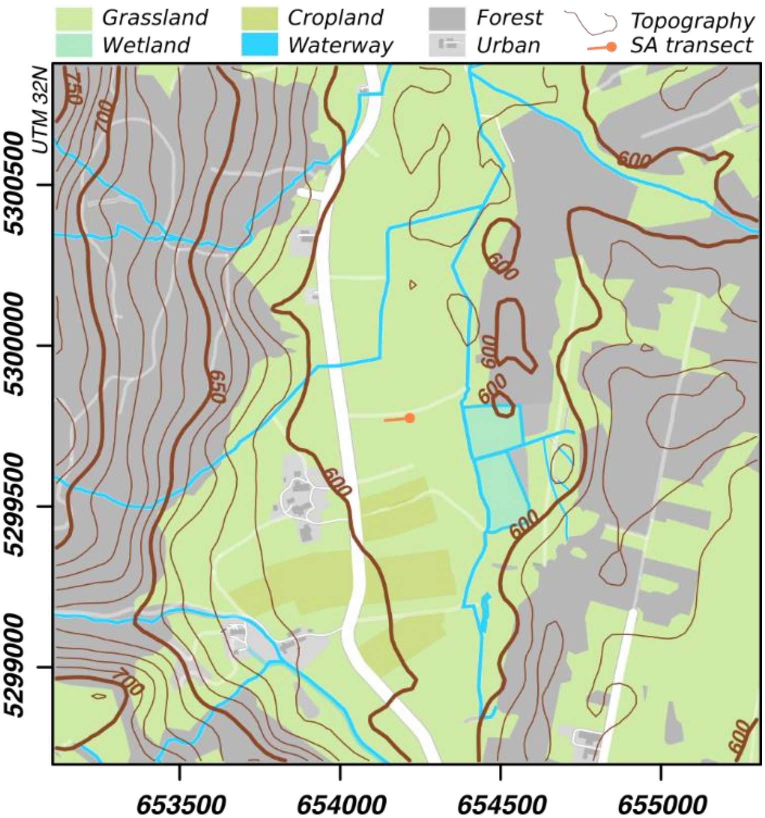


Figure 1: Location of the sonic anemometer (SA) transect at the DE-Fendt field. Map modified from Fig 1 in Zeeman et al. (2017)

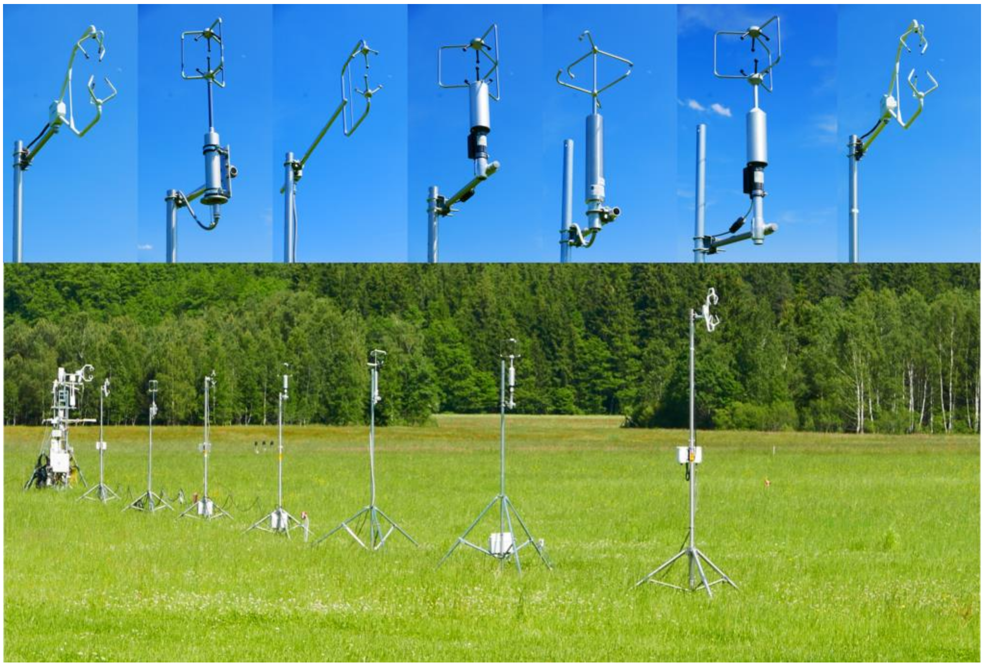
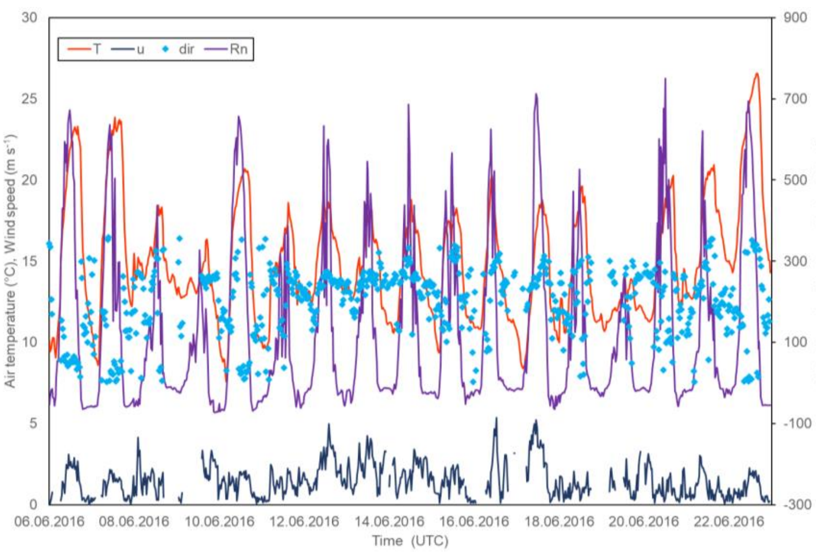


Figure 2: The bottom part of this figure shows a photograph of the field intercomparison experiment; the micrometeorological installations of the TERENO/ICOS site DE-Fen can be seen in the background (left). On top, close-up pictures of all seven sonic anemometers are show; they are presented from left to right in the same order as they are listed in Table 1.~~n.~~

Figure 3 shows the meteorological conditions during the experiment. As expected for this site and for this time of the year, the dominant daytime wind direction was North. Wind speeds ranged between 0 and 5 m s^{-1}. Air temperatures varied between 8° ° C and 24° C. Net radiation reached values up to 700 W m^{-2}. On 08, 09, 19 June, the cloud cover was rather dense all day. Most of the days are characterized be high loads of net radiation with values larger than 500 W m^{-2} at maximum. Nevertheless, also rain occurred on most of the days with the exception of first two days of the measurement period, 06 - 07 June, and the last day, 22 June. Overall, this experiment can be considered as being typical conditions in the early summer of temperate climate zones.

Table 1: Participating instruments in the order of their location from East to West

Comparison name	CSAT3_1	Gill.R3	Gill.HS	Metek.uSonic3.omni	Young.81000	Young.81000RE	CSAT3_2
Manufacturer	Campbell Scientific Inc.	Gill Instruments Ltd.	Gill Instruments Ltd.	METEK Meteorologische Messtechnik GmbH	R. M. Young Company	R. M. Young Company	Campbell Scientific Inc.
Model	CSAT3	1210R3	HS	uSonic-3 Omni AH	81000	81000RE	CSAT3
Serial number	1791	585	152903	0106054006	003149	UA 02043	0771
Path length (mm)	116	150	150	138	150	150	116
Transducer diameter (mm)	6.4	11	11	13.8	13.8	13.8	6.4
Transducer path angle (°)	60	45	45	45	45	45	60



5 **Figure 3: Meteorological elements during the intercomparison experiment; 30-min averages of air temperature (*T*) and net radiation (*Rn*) measured at 2 m, and wind speed (*u*) and direction (*dir*) from the DE-Fen site measured at 3.25 m.**

2.2 Data processing

All data were processed~~ing~~ using the TK3 software (Mauder and Foken, 2015) according to the processing scheme of Mauder et al. (2013). More precisely, turbulent statistics were calculated using 30-min block averaging, after applying a spike removal algorithm on the high-frequency raw data. We applied the double rotation method (Kaimal and Finnigan, 1994) and a spectral correction for path averaging according to Moore (Moore, 1986). The compared turbulent quantities are defined as follows:

$U = \bar{u}$, the averaged total wind velocity after alignment of the coordinate system into the mean wind (after double rotation);

$T_s = \overline{T}_s$, the averaged sonic temperature;

$\sigma_w = \sqrt{\overline{w'w'}}$, the standard deviation of the vertical velocity component;

$\sigma_{Ts} = \sqrt{\overline{T_s{}'T_s{}'}}$, the standard deviation of the sonic temperature;

$u_* = \sqrt[4]{\overline{u'w'}^2 + \overline{v'w'}^2}$, the friction velocity calculated from both covariances between the two horizontal wind components and w;

$H_s = \rho c_p \overline{w'T_s{}'}$, the buoyancy flux calculated from the air density ρ, the specific heat capacity at constant pressure c_p and the covariance between w and T_s.

These quantities were filtered for rain (during the respective half hour or the half hour before as recorded by a Vaisala WXT520 sensor of the nearby TERENO station DE-Fen), obstructed wind directions φ based on 30-minute averages (70° < φ < 110°; 250° < φ < 290°) and non-steady-state conditions, i.e. data with Foken et al. (2004) steady state test flag 4-9, considering the u_*-flag for all statistics concerning the pure wind measurements (U, σ_w, u_*) and the sensible heat flux-flag for all statistics that include sonic temperature (T_s, σ_{Ts}, H_s).

The reference instrument (etalon) was chosen for each compared quantity independently according a principle component analysis (PCA) using the R function princomp(). We selected the instrument with the highest loading on the first principle component. Only when the Young.81000RE had received the highest loading, we selected the sonic anemometer with the second highest loading as etalon instead because the Young.81000RE time series only starts more than three days later at 10.06.2016 14:00 due to technical issues in the beginning of the field experiment.

For the statistical analysis of the intercomparison, an orthogonal Deming regression was applied in order to account for measurement errors in both x and y variables, using the R package mcr (Manuilova, E., Schuetzenmeister and Model, 2014; R_Core_Team, 2016). Furthermore, we calculated the values for comparability, which is equivalent to the root-mean-square-error (RMSE), and bias, which is the mean error for a certain measurement quantity.

3 Results

3.1 Mean total wind ~~speed~~velocity

For our comparison of the mean wind velocity measurements, the METEK.uSonic3.omni was selected as etalon, because it received the highest loading (-0.3785) on the first principle component of our PCA. However, the loadings of the two Gill instruments and the YOUNG.81000 are not much lower either. Hence, the two Gill anemometers and the Young.81000 compare slightly better with the etalon than the rest. Nevertheless, the agreement between of the U-measurements by all tested anemometers is generally very good, as can be seen from Figure 4. This is also indicated by small regression intercepts (< 0.04 m s^{-1}) and slopes close to one (1±0.03). In general, comparability values are smaller than 0.11 m s^{-1} and biases range between −0.05 m s^{-1} and 0.06 m s^{-1} (Table 2). The agreement between the two CSAT3 is as good as the overall agreement between all tested instruments.

Table 2: Regression results for the comparison of mean total wind velocity *U*, plus estimates for bias and comparability (RMSE).

etalon = METEK.uSonic3.omni	CSAT3_1	Gill.R3	Gill.HS	Young.81000	Young.81000RE	CSAT3_2
n	367	367	367	366	257	367
intercept (m s^{-1})	-0.01	0.00	0.01	0.03	0.00	0.04
slope	1.02	1.02	1.03	1.02	0.97	0.97
bias (m s^{-1})	0.02	0.04	0.05	0.06	-0.05	0.00
RMSE (m s^{-1})	0.11	0.08	0.07	0.08	0.11	0.09

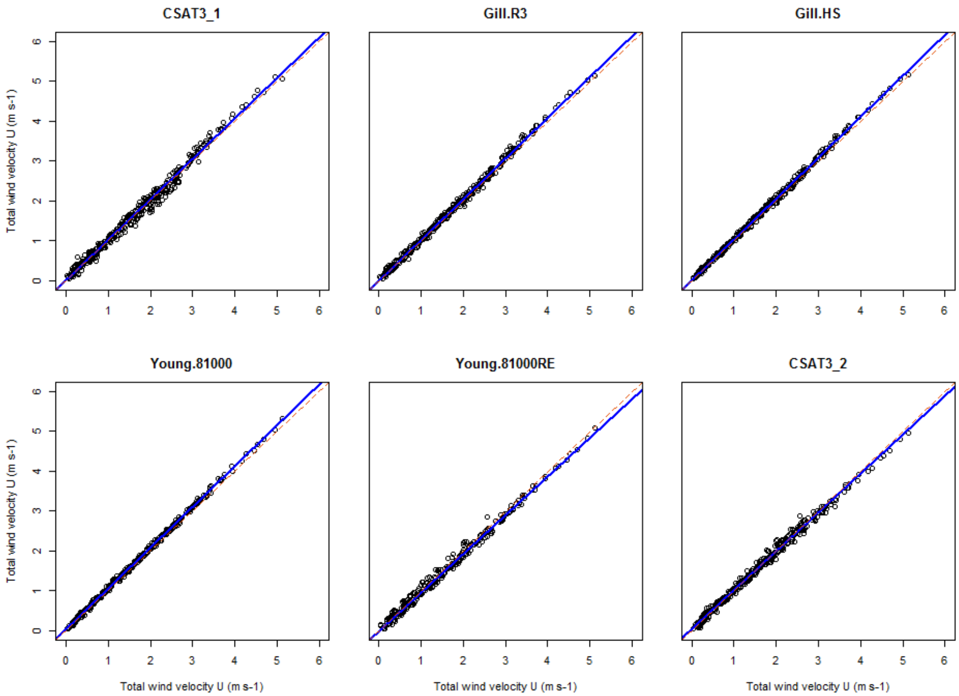


Figure 4: Comparison of the 30 min averaged total wind velocity measurements (etalon = METEK.uSonic.omni).

3.2 Mean sonic temperatures

The ultrasound-based temperature measurement is determined from the absolute time of flight as opposed to the differences in time flight for the velocity measurement. Therefore, inaccuracies in pathlength due to inadvertent bending or varying electronic delays of the signal processing directly affect the accuracy of the measurement, and it is not surprising that the general agreement between different instruments is much worse for the sonic temperature than for the wind velocity. The Young.81000 received the highest loading (-0.3806) and was therefore chosen as etalon. Good agreement with this reference is found for the two CSAT3 and the METEK.uSonic.omni, which is indicated by values well below 1 K for bias and comparability. However, larger discrepancies occur for the two Gill sonic anemometers and the Young.81000RE. As can be seen from Figure 5, the Young.81000RE sonic temperatures show a linear relationship with the etalon, so that the error of this instrument could be corrected by a simple regression equation using the coefficients provided in Table 3. In contrast, the sonic temperature measurements of the two Gill sensors show much more scatter and non-linearity in addition to a large bias, which is determined as 1.82 K for the Gill.R3 and 3.55 K for the Gill.HS. Therefore, the comparability values are also large with RMSE = 1.99 K for the Gill.R3 and 3.58 K for the Gill.HS.

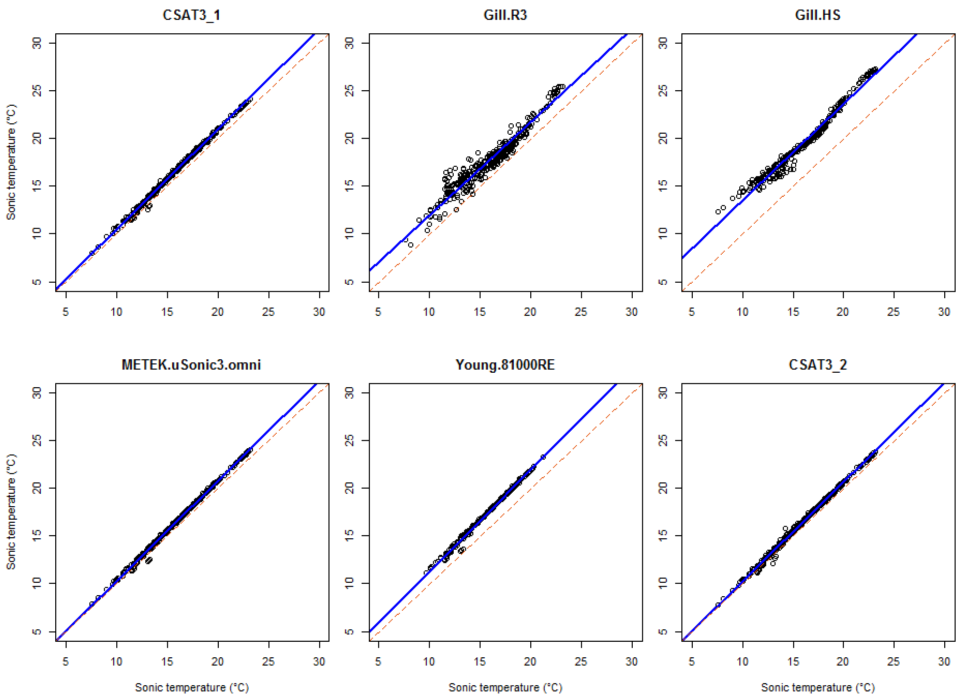


Figure 5: Comparison of the averaged sonic temperature measurements (etalon = Young.81000).

Table 3: Regression results for the comparison of mean sonic temperature T_s, plus estimates for bias and comparability (RMSE); unusually large deviations from the etalon are underlined (slopes deviating more than 5% from unity and absolute differences of more than 1 K).

etalon = Young.81000	CSAT3_1	Gill.R3	Gill.HS	METEK.uSonic3.omni	Young.81000RE	CSAT3_2
n	321	321	321	321	229	321
intercept (K)	0.01	2.22	3.37	-0.18	0.69	-0.15
slope	1.05	0.97	1.01	1.05	1.06	1.04
bias (K)	0.75	1.82	3.55	0.58	1.69	0.49
RMSE (K)	0.79	1.99	3.58	0.62	1.70	0.54

3.3 Standard deviation of the vertical velocity component

An accurate and precise measurement of the standard deviation of the vertical velocity component is particularly important because the *w*-fluctuations are required for the determination of any scalar flux by eddy covariance, also those fluxes that require the deployment of an additional sensor, such as an infrared gas analyser or other laser-based fast-response sensors. During our field experiment, σ_w values ranged between 0 m s^{-1} and 0.7 m s^{-1}. The Gill.HS anemometer was chosen as etalon for σ_w as it received the highest loading from our PCA (-0.3781). All other instruments agree very well with this reference as can be seen from Figure 6. Intercepts and biases are very small, ranging from −0.01 m s^{-1} to 0.02 m s^{-1} (Table 4). Values for comparability are better than 0.02 m s^{-1} and the regression slopes are close to one (1±0.03).

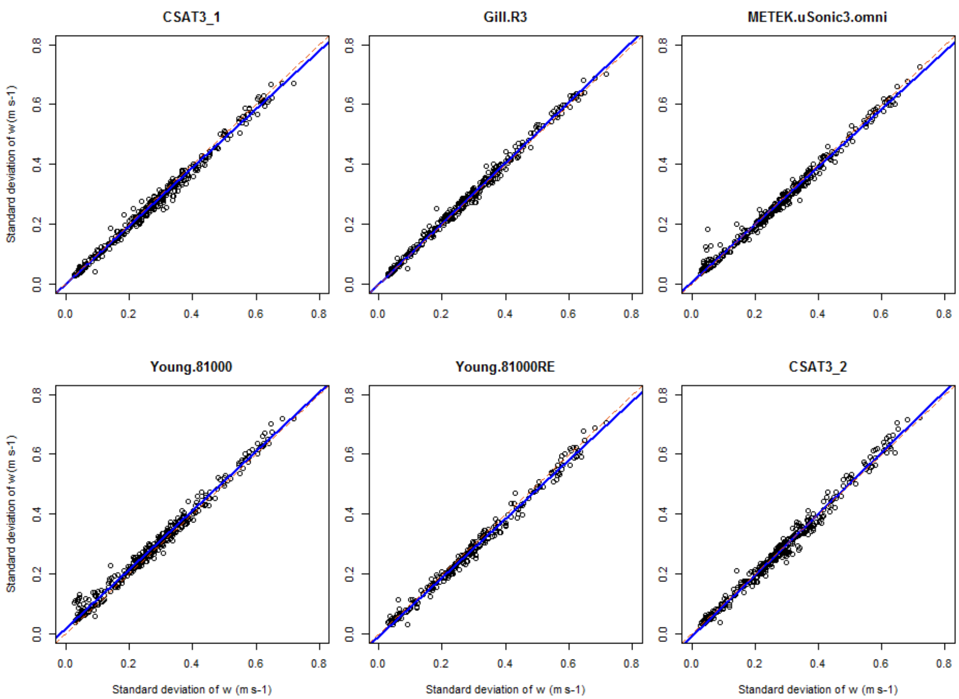


Figure 6: Comparison of the standard deviation of the vertical velocity component σ_w (etalon = Gill.HS)

Table 4: Regression results for the comparison of the standard deviation σ_w, plus estimates for bias and comparability (RMSE).

etalon = Gill.HS	CSAT3_1	Gill.R3	METEK.uSonic3.omni	Young.81000	Young.81000RE	CSAT3_2
n	367	367	367	366	257	367
intercept (m s^{-1})	0.00	0.00	0.00	0.02	-0.01	-0.01
slope	0.98	1.01	0.97	0.99	0.98	1.03
bias (m s^{-1})	-0.01	0.00	0.00	0.01	-0.01	0.00
RMSE (m s^{-1})	0.02	0.01	0.02	0.02	0.02	0.02

3.4 Standard deviation of the sonic temperature

Despite the large discrepancies of the mean sonic temperature measurements ~~between certain~~ of the Gill instruments, the fluctuations of sonic temperature agree much better (Figure 7). For this turbulent quantity, the CSAT2_2 was chosen as etalon, although it only had the second-highest loading in our PCA (-0.3816) because the Young.81000RE, which received a slightly higher loading (-0.3824), only recorded data four days after the comparison experiment had begun. None of the tested instruments shows a large bias nor a large regression intercept for the measurement of σ_{Ts}. However, the large errors in mean sonic temperature of the two Gill anemometers also lead to a larger scatter for σ_{Ts}, which expresses itself in comparability values larger than 0.06 K for the Gill.HS and 0.08 K for the Gill.R3 (Table 5). Surprisingly, the Young.81000 has an even poorer comparability of 0.09 K – it was the etalon for the mean sonic temperature measurement. In contrast, the Young.81000RE shows a very good agreement with the etalon for σ_{Ts} despite its large bias when measuring mean sonic temperature. The METEK.uSonic.omni stands out because it has the highest regression slope of 1.06, which ~~might be~~is a direct consequence of the almost equally high regression slope of 1.05 for the mean sonic temperature measurement. The agreement between the two CSAT3 is very good except for a few outliers, which were not rejected by our data-screening algorithm.

Table 5: Regression results for the comparison of the standard deviation σ_{Ts}, plus estimates for bias and comparability (RMSE); unusually large deviations from the etalon are underlined (slopes deviating more than 5% from unity and absolute differences larger than 0.05 K).

etalon = CSAT3_2	CSAT3_1	Gill.R3	Gill.HS	METEK.uSonic3.omni	Young.81000	Young.81000RE
n	322	322	322	322	321	229
intercept	0.00	0.01	0.01	-0.01	0.02	0.00
slope	1.01	0.99	0.96	<u>1.06</u>	1.05	1.04
bias (K)	0.00	0.00	-0.01	0.01	0.04	0.01
RMSE (K)	0.05	<u>0.08</u>	<u>0.06</u>	0.05	<u>0.09</u>	0.03

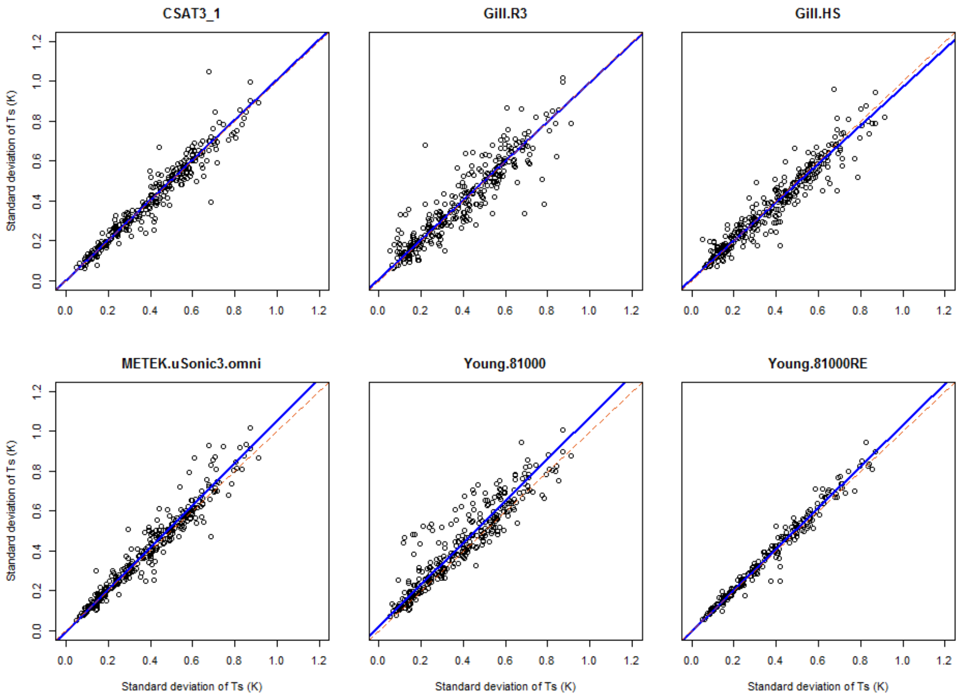


Figure 7: Comparison of the standard deviation of the sonic temperature σ_{Ts} (etalon = CSAT3_2).

3.5 Friction velocity

Friction velocities ranged between 0 m s^{-1} and almost 0.6 m s^{-1} during our experiment. Although the Youn~~t~~g.81000RE has the highest loading (-0.3803) in our PCA, we chose the Gill.HS as etalon due to the abovementioned data-availability issue of the Young.81000RE, but again its loading is only slightly lower (-0.3801). For u_*, generally much larger scatter is observed than for other purely wind-related quantities, such as U and σ_w (Figure 8), which manifests itself in comparability values of 0.05 m s^{-1} or 0.06 m s^{-1} respectively (Table 6). However, despite the large scatter the biases and regression intercepts are generally smaller with values lower than 0.02 m s^{-1} in absolute numbers. Only the METEK.uSonic.omni measures friction velocities consistently larger than the etalon on average, which manifests itself in a bias and regression intercept of 0.03 m s^{-1}. The relatively low regression slope of the CSAT3_1 of 0.91 does not lead to unusually poor error estimates of neither comparability (0.05 m s^{-1}) nor bias (−0.01 m s^{-1}). Similarly, the CSAT_2 shows the second lowest regression slope, but its bias and RMSD is very similar to the other instruments.

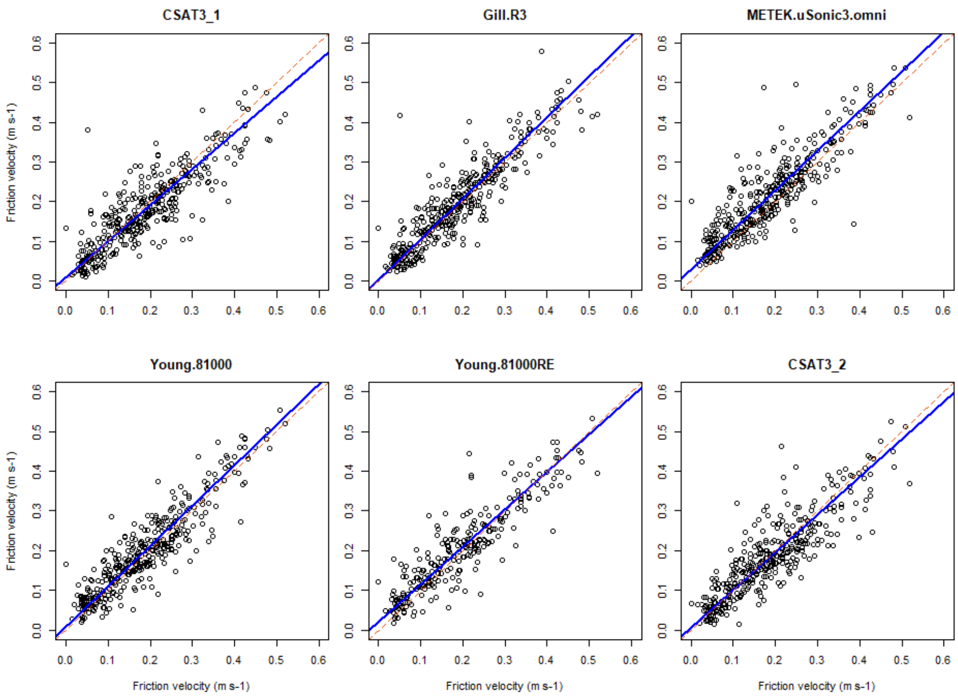


Figure 8: Comparison of the friction velocity measurements (etalon= Gill.HS).

5

Table 6: Regression results for the comparison of friction velocity u_*, plus estimates for bias and comparability (RMSE); unusually large deviations from the etalon are underlined (slopes deviating more than 5% from unity).

etalon = Gill.HS	CSAT3_1	Gill.R3	METEK.uSonic3.omni	Young.81000	Young.81000RE	CSAT3_2
n	365	362	365	364	255	364
intercept (m s^{-1})	0.01	0.00	0.03	0.01	0.02	0.01
slope	<u>0.91</u>	1.03	1.00	1.02	0.95	0.95
bias (m s^{-1})	-0.01	0.01	0.03	0.01	0.01	0.00
RMSE (m s^{-1})	0.05	0.05	0.06	0.05	0.05	0.05

3.6 Buoyancy flux

Quantifying fluxes by eddy covariance is probably the most common application of sonic anemometers. Therefore, the comparison of the buoyancy flux measurements is perhaps the most interesting aspect of this study ~~from~~ for many researchers. At first, we would like to note that the number of available data is reduced by about one third compared to the other quantities, which is due to rejection of instationary periods by the quality tests of Foken et al. (2004). ~~For this comparison, t~~The CSAT3_1 was chosen as etalon for this quantity because it received the highest loading in our PCA (-0.3786). The overall agreement between all sonic anemometers is excellent as can be seen from Figure 9. Biases are generally very small with values less than 3 W m^{-2}, and all of the regression slopes are very close to one (1±0.02) (Table 7). Some minor scatter that is apparent in the comparison plots of Figure 9 results in comparability values between 8.6 W m^{-2} and 11.2 W m^{-2} for the different instruments.

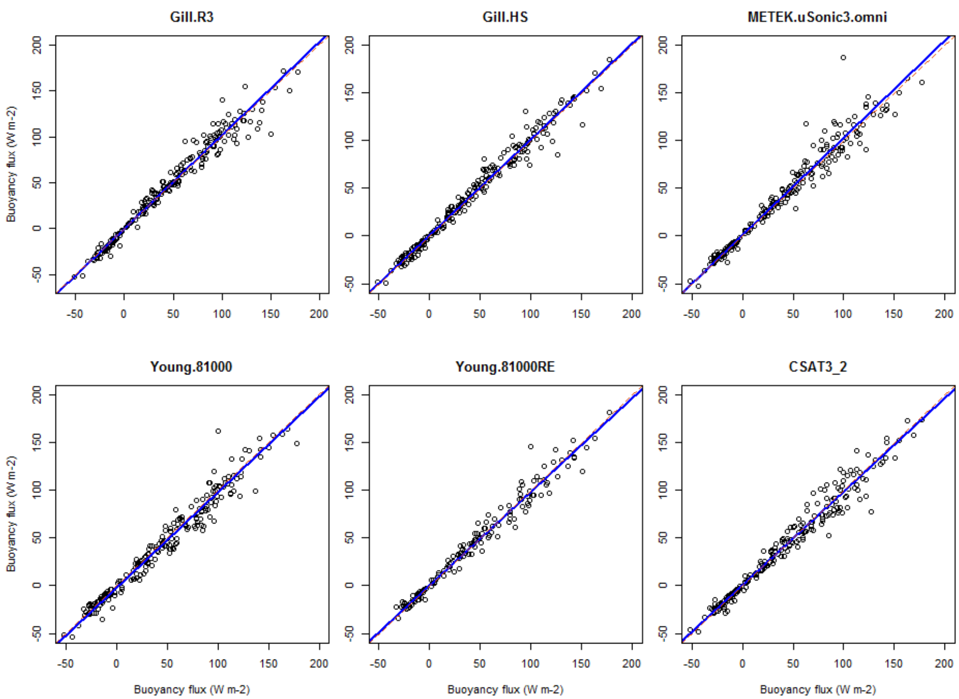


Figure 9: Comparison of buoyancy flux measurements (etalon = CSAT3_1)

Table 7: Regression results for the comparison of buoyancy flux H_s, plus estimates for bias and comparability (RMSE)

etalon = CSAT3_1	Gill.R3	Gill.HS	METEK.uSonic3.omni	Young.81000	Young.81000RE	CSAT3_2
n	219	224	209	210	153	211
intercept (W m^{-2})	0.0	1.2	0.9	-2.5	0.8	0.7
slope	1.02	1.00	1.02	1.00	0.98	0.98
bias (W m^{-2})	0.7	1.4	1.7	-2.6	0.0	-0.3
RMSE (W m^{-2})	9.4	8.6	11.2	10.5	8.6	10.0

4 Discussion

In theory, the overall agreement between sonic anemometers cannot be better than the random error, if the seven different measurement~~s~~ systems collect independent samples of an homogeneous turbulence field (Richardson et al., 2012). The stochastic error ~~of~~ due to limited sampling of the turbulent ensemble (Finkelstein and Sims, 2001) is 17% or 0.03 m s^{-1} on average for u_* and 14% or 5 W m^{-2} for H_s, based on data from CSAT3_1. The comparability values that we found between different instruments for these two quantities are only slightly larger. This means, a better agreement is hardly physically possible, and the remaining small discrepancies can be explained by slight surface heterogeneities within the footprint area of the different systems and by a very small instrumental error. The agreement between the two CSAT3 ~~was~~ is as good as the agreement with other sonic anemometer models. ~~The rainguards on the CSAT3 unit with serial number 1791 had no visible influence on the measurement performance in comparison to the unit with serial number 0771, and the number of available half-hour statistics was not affected either.~~

We found a much better agreement between different sonic anemometers, especially for u_* and H_s, in comparison to previous intercomparison experiments (Loescher et al., 2005; Mauder et al., 2007). Now, all tested instruments are within the limits that Mauder et al. (~~Mauder et al.,~~ 2006) classified as type A, i.e. sonic anemometers suitable for fundamental turbulence research. Perhaps this can partially be explained by a consistent digital data acquisition, implemented here with a very high precision clock and event-driven communication using Python programming language. Probably, the implementation of a more efficient spike removal algorithm for the high-frequency data and other additional quality tests in the post-processing scheme of Mauder et al. (2013) also helped to improve the data quality of the resulting fluxes and consequently improved the agreement. ~~A contribution by changes in the firmware of the different sonic anemometers over the last ten years are likely but not fully documented.~~

On top of that, the filtering for obstructed wind direction sectors and for rain, as described in section 2.2, was crucial to remove poor quality data. Both additional steps improved the agreement between instruments considerably. For σ_w, regression slopes ranged between 1.00 and 1.24 and intercepts were between −0.05 and 0.00 m s^{-1} after processing according to Mauder et al. (2013). After filtering for obstructed wind direction, slopes ranged between 0.98 and 1.22 and intercepts remained between

Formatiert: Schriftart: Kursiv

Formatiert: Schriftart: Kursiv, Tiefgestellt

−0.05 and 0.00 m s^{-1}. As can be seen from the results (Table 4), the overall agreement further improved after the filtering for rainy periods. Especially, some outliers of the CSAT3_2, which did not have the rain-guard meshes at the transducer heads, were rejected after this step. The effect of the data filtering on other quantities, such as Hs, was smaller. Here, the slopes ranged already only between 0.97 and 1.00 after processing according to Mauder et al. (2013), which did not change much further after filtering for obstructed wind directions and for rainy periods (Table 7). This can be explained by the fact that the scheme of Mauder et al. (2013) is designed for quality control of fluxes and not necessarily standard deviations. It therefore much stricter for H_s than for σ_w.

~~Especially c~~Considering flow distortion errors on the order of 5% or more that are reported in the literature (Frank et al., 2016; Horst et al., 2015; Huq et al., 2017), the very good agreement between all sonic anemometers in this field experiment is nevertheless somewhat surprising. A contribution by changes in the firmware of the different sonic anemometers over the last ten years are likely but not fully documented. According to the manufacturer, the two CSAT3 sonic anemometers have no flow distortion correction at all, while all the other five instruments probably do apply some sort of correction, only the exact details are not ~~known~~ publically available for all of them. This could mean that flow distortion errors are indeed significant for our experiment but perhaps all instruments are afflicted with an error of almost the exact same magnitude and consequently underestimate σ_w and vertical scalar fluxes similarly, despite the obvious differences in sensor geometry and internal data processing.

Alternatively, one might also suppose that the flow distortion errors were generally small for our experimental set-up due to the occurred distribution of instantaneous flow angles, since flow-distortion effects tend to be smaller for smaller angles of attack as indicated by the studies of Grelle and Lindroth (1994) and Gash and Dolman (2003). However, the standard deviation of the angles-of-attack was about 15°, which is comparable to other field experiments. For comparison, Gash and Dolman (2003) report about 90% of their data to be within ±20° for the Horstermeer peat bog site, and Grare et al. (2016) report their data to be in a range of ±15°, most of times even within ±10°, measuring at 10 m above shrubland. Horst et al. (2015) report their angles-of-attack to be mostly within ±8° for measurements above low weeds and crop stubble with an aerodynamic roughness length of 0.02 m. Since the spread of angles of attack is on the upper end of the values reported in the literature, our comparison results can be considered as a conservative estimate for the random instrument-related uncertainty of typical applications of eddy-covariance measurements over vegetation canopies. A common significant systematic error of all tested instruments ~~might nevertheless be~~is quite possible, as suggested by Frank et al. (2016).

One exception from the overall very good agreement is the sonic temperature measurement by both Gill sonic anemometers, the HS and the R3. This error appears not only as an offset, but also as deviation of a linear functional relationship and increased scatter. A similar behaviour of other Gill anemometers has been reported before, and also a possibly explanation has been provided in the past (Mauder et al., 2007; Vogt, 1995). Obviously, the sonic temperature measurement of Gill anemometers is compromised by a temperature-dependence of the transducer delay, i.e. the time delay between the arrival of a sound pulse at the transducer and the registration by the electronics board.

Formatiert: Hochgestellt

Formatiert: Schriftart: Kursiv

Formatiert: Schriftart: Kursiv, Tiefgestellt

Formatiert: Schriftart: Kursiv

Formatiert: Schriftart: Kursiv, Tiefgestellt

5 Conclusions

Generally, biases and regression intercepts were very small for all sensors and all computed variables, except for the temperature measurements of the two Gill sonic anemometers (HS and R3), which are known to have a transducer-temperature dependence of the sonic temperature measurement (Mauder et al., 2007). Nevertheless, the Gill anemometers show an equally good agreement for other turbulence statistics. The comparability (RMSE) of the instruments is not always as good as the bias, indicating a random error that is slightly larger than any systematic discrepancies. The best overall agreement between the different instruments was found for the quantities U , σ_w, and H_s, which suggests that the sensors' physical structure and internal signal processing are designed for measuring wind speed and vertical scalar fluxes as accurately as possible. However, the relative random uncertainty of u_* measurements is still large, pointing at the particular challenge in measuring the covariance of horizontal and vertical wind components due to the rather small spectral overlap.

The uncertainty estimate of Mauder et al. (2006) for the buoyancy flux measurement of 5% or 10 W m^{-2} was confirmed, not only for those instruments that were classified in that study as "type A" (CSAT3 and Gill HS), but also for those that were labelled "type B" (Gill R3) back then and all other tested instruments (METEK uSonic3-omni, RM Young 81000 and 81000RE). ~~The uncertainty estimate of Mauder et al. (2006) for the buoyancy flux measurement of 5% or 10 W m^−2 was confirmed, not only for those instruments that were classified in that study as "type A" (CSAT3 and Gill HS), but also for those that were labelled "type B" (Gill R3) back then and all other tested instruments (METEK uSonic3-omni, RM Young 81000 and 81000RE).~~ Hence, from our results we cannot derive a classification of the tested sonic anemometers in different quality levels, which means that the evolution of anemometers by all major manufacturers has converged over the last decade. For applications aiming at measuring vertical scalar fluxes, all tested instruments can be considered equally suitable~~, at least for low vegetation ecosystems~~, as long as digital data acquisition is implemented to avoid additional uncertainty and a stringent data quality control procedure is applied to detect malfunction of the eddy-covariance system. Moreover, the deviations between instruments of different manufacturers are not larger than between different serial numbers of the same model. Therefore, we do not consider it to be necessary to agree on one single anemometer model to ensure comparability, e.g. for intensive field campaigns or for networks of ecosystem observatories. Instead, other criteria should be taken into account for the selection of a sonic anemometer, such as climatic conditions of a measurement site (e.g. frost, fog, heat), the distribution of wind directions (omnidirectional or not), the measurement height (path length), the compatibility with an existing data acquisition system or a certain scientific objective. In principle, this conclusion is not in contradiction with the classification Foken and Oncley (1995) and Mauder et al. (2006), because they also concluded that all instruments under investigation were suitable for general flux measurements. Only for specific questions of fundamental turbulence research, it was advised to use ~~specific~~ certain types of instruments.

Although a good agreement between six different sonic anemometer models indicates a high precision of these type of instruments in general, a field intercomparison study can only provide limited insights on the absolute accuracy of these measurements. Particularly, a systematic error that is common to all tested instruments can inherently never be detected in this way. In the past, wind-tunnel experiments were conducted for this purpose, although their transferability to real-world conditions was always debated. Numerical simulations of probe-induced flow distortion (Huq et al., 2017) may provide a better way to characterize the suitability of sonic anemometers for turbulence measurements in the future. If systematic errors for one certain instrument are known from these computationally very expensive simulations, then classical field inter-comparisons can be used to test models against such a well-characterized sensor. Moreover, a comparison with a remote sensing based system that is free of flow-distortion, such as LiDAR, would be very helpful if it is able to sample a similarly small volume of air at a similar measurement rate as a sonic anemometer.

Acknowledgements

We are grateful to Kevin Wolz and Peter Brugger for field assistance and to Sandra Genzel for help with the data preparation. We thank GWU-Umwelttechnik GmbH for generously lending us the Young.81000RE instrument for the duration of the field experiment. The Fendt site is part of the TERENO and ICOS-D ecosystems (Integrated Carbon Observation System, Germany) networks which are funded, in part, by the German Helmholtz Association and the German Federal Ministry of Education and Research (BMBF). We thank the Scientific Team of ScaleX Campaign 2016 for their contribution. This work was conducted within the Helmholtz Young Investigator Group "Capturing all relevant scales of biosphere-atmosphere exchange – the enigmatic energy balance closure problem", which is funded by the Helmholtz-Association through the President's Initiative and Networking Fund.

References

Burns, S. P., Horst, T. W., Jacobsen, L., Blanken, P. D. and Monson, R. K.: Using sonic anemometer temperature to measure sensible heat flux in strong winds, Atmos.Meas.Tech., 5(9), 2095–2111, doi:10.5194/amt-5-2095-2012, 2012.

Dyer, A. J.: Flow distorsion by supporting structures, Boundary-Layer Meteorol., 20, 243–251, 1981.

Dyer, A. J., Garratt, J. R., Francey, R. J., McIlroy, I. C., Bacon, N. E., Bradley, E. F., Denmead, O. T., Tsvang, L. R., Volkov, Y. A., Koprov, B. M., Elagina, L. G., Sahashi, K., Monji, N., Hanafusa, T., Tsukamoto, O., Frenzen, P., Hicks, B. B., Wesely, M., Miyake, M., Shaw, W., Hyson, P., McIlroy, I. C., Bacon, N. E., Victoria, A., Bradley, E. F., Tsvang, L. R., Volkov, Y. A., Koprov, B. M., Elagina, L. G., Sahashi, K., Monji, N., Hanafusa, T., Hicks, B. B., Frenzen, P., Wesely, M., Miyake, M. and Shaw, W.: An international turbulence comparison experiment (ITCE-76), Boundary-Layer Meteorol., 24, 181–209, 1982.

Finkelstein, P. L. and Sims, P. F.: Sampling error in eddy correlation flux measurements, J. Geophys. Res., 106(D4), 3503–3509, doi:10.1029/2000JD900731, 2001.

Foken, T. and Oncley, S. P.: Workshop on instrumental and methodical problems of land surface flux measurements, Bull. Am. Meteorol. Soc., 76, 1191–1193, 1995.

Foken, T., Göckede, M., Mauder, M., Mahrt, L., Amiro, B. and Munger, W.: Post-field data quality control, in Handbook of Micrometeorology. A Guide for Surface Flux Measurement and Analysis, edited by X. Lee, W. Massman, and Law. Beverly,

pp. 181–208, Kluwer Academic Publishers, Dordrecht., 2004.
Frank, J. M., Massman, W. J. and Ewers, B. E.: Underestimates of sensible heat flux due to vertical velocity measurement errors in non-orthogonal sonic anemometers, Agric. For. Meteorol., 171–172, 72–81, doi:10.1016/j.agrformet.2012.11.005, 2013.
Frank, J. M., Massman, W. J., Swiatek, E., Zimmerman, H. A. and Ewers, B. E.: All sonic anemometers need to correct for transducer and structural shadowing in their velocity measurements, J. Atmos. Ocean. Technol., 33(1), 149–167, doi:10.1175/JTECH-D-15-0171.1, 2016.
Gash, J. H. C. and Dolman, A. J.: Sonic anemometer (co)sine response and flux measurement: I. The potential for (co)sine error to affect sonic anemometer-based flux measurements, Agric. For. Meteorol., 119(3–4), 195–207, doi:10.1016/S0168-1923(03)00137-0, 2003.
Grelle, A. and Lindroth, A.: Flow Distortion by a Solent Sonic Anemometer: Wind Tunnel Calibration and Its Assessment for Flux Measurements over Forest and Field, J. Atmos. Ocean. Technol., 11(6), 1529–1542, doi:10.1175/1520-0426(1994)011<1529:FDBASS>2.0.CO;2, 1994.
Högström, U.: A critical evaluation of the aerodynamical error of a turbulence instrument, J. Appl. Meteorol., 21, 1838–1844, 1982.
Högström, U. and Smedman, A. S.: Accuracy of sonic anemometers: Laminar wind-tunnel calibrations compared to atmospheric in situ calibrations against a reference instrument, Boundary-Layer Meteorol., 111(1), 33–54, doi:10.1023/B:BOUN.0000011000.05248.47, 2004.
Horst, T. W., Semmer, S. R. and Maclean, G.: Correction of a Non-orthogonal, Three-Component Sonic Anemometer for Flow Distortion by Transducer Shadowing, Boundary-Layer Meteorol., 155(3), 371–395, doi:10.1007/s10546-015-0010-3, 2015.
Huq, S., De Roo, F., Foken, T. and Mauder, M.: Evaluation of probe-induced flow distortion of Campbell CSAT3 sonic anemometers by numerical simulation, Boundary-Layer Meteorol., in press, doi:10.1007/s10546-017-0264-z, 2017.
Kaimal, J. .: Sonic Anemometer Measurement of Atmospheric Turbulence, in Proceedings of the Dynamic Flow Conference 1978 on Dynamic Measurements in Unsteady Flows, edited by B. W. Hanson, pp. 551–565, Springer Netherlands., 1979.
Kaimal, J. C. and Finnigan, J. J.: Atmospheric Boundary Layer Flows: Their Structure and Measurement, Oxford University Press, New York, NY., 1994.
Kochendorfer, J., Meyers, T. P., Heuer, M. W., Frank, J. M., Massman, W. J. and Heuer, M. W.: How well can we measure the vertical wind speed? Implications for the fluxes of energy and mass, Bound. Layer Meteorol., 145, 383–398, doi:10.1007/s10546-012-9738-1, 2012.
Kochendorfer, J., Meyers, T. P., Frank, J. M., Massman, W. J. and Heuer, M. W.: Reply to the Comment by Mauder on "How Well Can We Measure the Vertical Wind Speed? Implications for Fluxes of Energy and Mass," Boundary-Layer Meteorol., 147(2), 337–345, doi:10.1007/s10546-012-9792-8, 2013.
Loescher, H. W., Ocheltree, T., Tanner, B., Swiatek, E., Dano, B., Wong, J., Zimmerman, G., Campbell, J., Stock, C., Jacobsen, L., Shiga, Y., Kollas, J., Liburdy, J. and Law, B. E.: Comparison of temperature and wind statistics in contrasting environments

among different sonic anemometer-thermometers, Agric. For. Meteorol., 133(1–4), 119–139, doi:10.1016/j.agrformet.2005.08.009, 2005.

Lugauer, M. and Winkler, P.: Thermal circulation in South Bavaria – climatology and synoptic aspects, Meteorol. Zeitschrift, 14(1), 15–30, doi:10.1127/0941-2948/2005/0014-0015, 2005.

Manuilova, E., Schuetzenmeister, A. and Model, F.: mcr: Method Comparison Regression, [online] Available from: https://cran.r-project.org/package=mcr, 2014.

Mauder, M.: A comment on “How well can we measure the vertical wind speed? Implications for fluxes of energy and mass” by Kochendorfer et al., Boundary-Layer Meteorol., 147(2), 329–335, doi:10.1007/s10546-012-9794-6, 2013.

Mauder, M., Liebethal, C., Göckede, M., Leps, J. P., Beyrich, F. and Foken, T.: Processing and quality control of flux data during LITFASS-2003, Boundary-Layer Meteorol., 121(1), 67–88, doi:10.1007/s10546-006-9094-0, 2006.

Mauder, M., Oncley, S. P., Vogt, R., Weidinger, T., Ribeiro, L., Bernhofer, C., Foken, T., Kohsiek, W., Bruin, H. a R. and Liu, H.: The energy balance experiment EBEX-2000. Part II: Intercomparison of eddy-covariance sensors and post-field data processing methods, Boundary-Layer Meteorol., 123(1), 29–54, doi:10.1007/s10546-006-9139-4, 2007.

Miyake, M., Stewart, R. W., Burling, H. W., Tsvang, L. R., Koprov, B. M. and Kuznetsov, O. A.: Comparison of acoustic instruments in an atmospheric turbulent flow over water, Boundary-Layer Meteorol., 2, 228–245, 1971.

van der Molen, M. K., Gash, J. H. C. and Elbers, J. A.: Sonic anemometer (co)sine response and flux measurement II. The effect of introducing an angle of attack dependent calibration, Agric. For. Meteorol., 122(1–2), 95–109, doi:10.1016/j.agrformet.2003.09.003, 2004.

Moore, C. J.: Frequency response corrections for eddy correlation systems, Boundary-Layer Meteorol., 37(1–2), 17–35, doi:10.1007/BF00122754, 1986.

Nakai, T. and Shimoyama, K.: Ultrasonic anemometer angle of attack errors under turbulent conditions, Agric. For. Meteorol., 162–163, 14–26, doi:10.1016/j.agrformet.2012.04.004, 2012.

Nakai, T., van der Molen, M. K., Gash, J. H. C. and Kodama, Y.: Correction of sonic anemometer angle of attack errors, Agric. For. Meteorol., 136, 19–30, doi:10.1016/j.agrformet.2006.01.006, 2006.

R_Core_Team: R: A language and environment for statistical computing, [online] Available from: https://www.r-project.org/, 2016.

Richardson, A. D., Aubinet, M., Barr, A. G., Hollinger, D. Y., Ibrom, A., Lasslop, G. and Reichstein, M.: Uncertainty quantification, in Eddy Covariance: A Practical Guide to Measurement and Data Analysis, edited by M. Aubinet, T. Vesala, and D. Papale, pp. 173–210, Springer, Dordrecht., 2012.

Stoy, P. C., Mauder, M., Foken, T., Marcolla, B., Boegh, E., Ibrom, A., Arain, M. A., Arneth, A., Aurela, M., Bernhofer, C., Cescatti, A., Dellwik, E., Duce, P., Gianelle, D., van Gorsel, E., Kiely, G., Knohl, A., Margolis, H., Mccaughey, H., Merbold, L., Montagnani, L., Papale, D., Reichstein, M., Saunders, M., Serrano-Ortiz, P., Sottocornola, M., Spano, D., Vaccari, F. and Varlagin, A.: A data-driven analysis of energy balance closure across FLUXNET research sites: The role of landscape-scale heterogeneity, Agric. For. Meteorol., 171–172, 137–152, doi:10.1016/j.agrformet.2012.11.004, 2013.

Tsvang, L. R., Koprov, B. M., Zubkovskii, S. L., Dyer, A. J., Hicks, B., Miyake, M., Stewart, R. W. and McDonald, J. W.: A comparison of turbulence measurements by different instruments; Tsimlyansk field experiment 1970, Boundary-Layer Meteorol., 3, 499–521, 1973.

Tsvang, L. R., Zubkovskij, S. L., Kader, B. A., Kallistratova, M. A., Foken, T., Gerstmann, W., Przandka, Z., Pretel, J., Zelenny, J. and Keder, J.: International turbulence comparison experiment (ITCE-81), Boundary-Layer Meteorol., 31, 325–348, 1985.

Vogt, R.: Theorie, Technik und Analyse der experimentellen Flussbestimmung am Beispiel des Hartheimer Kiefernwaldes, Wepf, Basel., 1995.

Formatiert: Deutsch (Deutschland)

Wieser, A., Fiedler, F. and Corsmeier, U.: The influence of the sensor design on wind measurements with sonic anemometer systems, J. Atmos. Ocean. Technol., 18(10), 1585–1608, doi:10.1175/1520-0426(2001)018<1585:TIOTSD>2.0.CO;2, 2001.

Wolf, B., Chwala, C., Fersch, B., Gravelmann, J., Junkermann, W., Zeeman, M. J., Angerer, A., Adler, B., Beck, C., Brosy, C., Brugger, P., Emeis, S., Dannenmann, M., De Roo, F., Diaz-Pines, E., Haas, E., Hagen, M., Hajsek, I., Jacobeit, J., Jagdhuber, T., Kalthoff, N., Kiese, R., Kunstmann, H., Kosak, O., Krieg, R., Malchow, C., Mauder, M., Merz, R., Notarnicola, C., Philipp, A., Reif, W., Reineke, S., Rödiger, T., Ruehr, N., Schäfer, K., Schrön, M., Senatore, A., Shupe, H., Völksch, I., Wanninger, C., Zacharias, S. and Schmid, H. P.: The ScaleX campaign: scale-crossing land-surface and boundary layer processes in the TERENO-preAlpine observatory, Bull. Am. Meteorol. Soc., in press, doi:10.1175/BAMS-D-15-00277.1, 2017.

Zeeman, M. J., Mauder, M., Steinbrecher, R., Heidbach, K., Eckart, E. and Schmid, H. P.: Reduced snow cover affects productivity of upland temperate grasslands, Agric. For. Meteorol., 232, 514–526, doi:10.1016/j.agrformet.2016.09.002, 2017.

Zhang, S. F., Wyngaard, J. C., Businger, J. A. and Oncley, S. P.: Response characteristics of the U.W. sonic anemometer, J. Atmos. Ocean. Technol., 3, 315–323, doi:10.1175/1520-0426(1986)003<0315:RCOTUS>2.0.CO;2, 1986.