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

5   *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

10   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

15   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

20   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

25   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

30   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

5   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

10 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

15 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.

20 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

25 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

30 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).

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

5   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

10   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

15   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,

20   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

25   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

30   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

5   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. 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.

10   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.
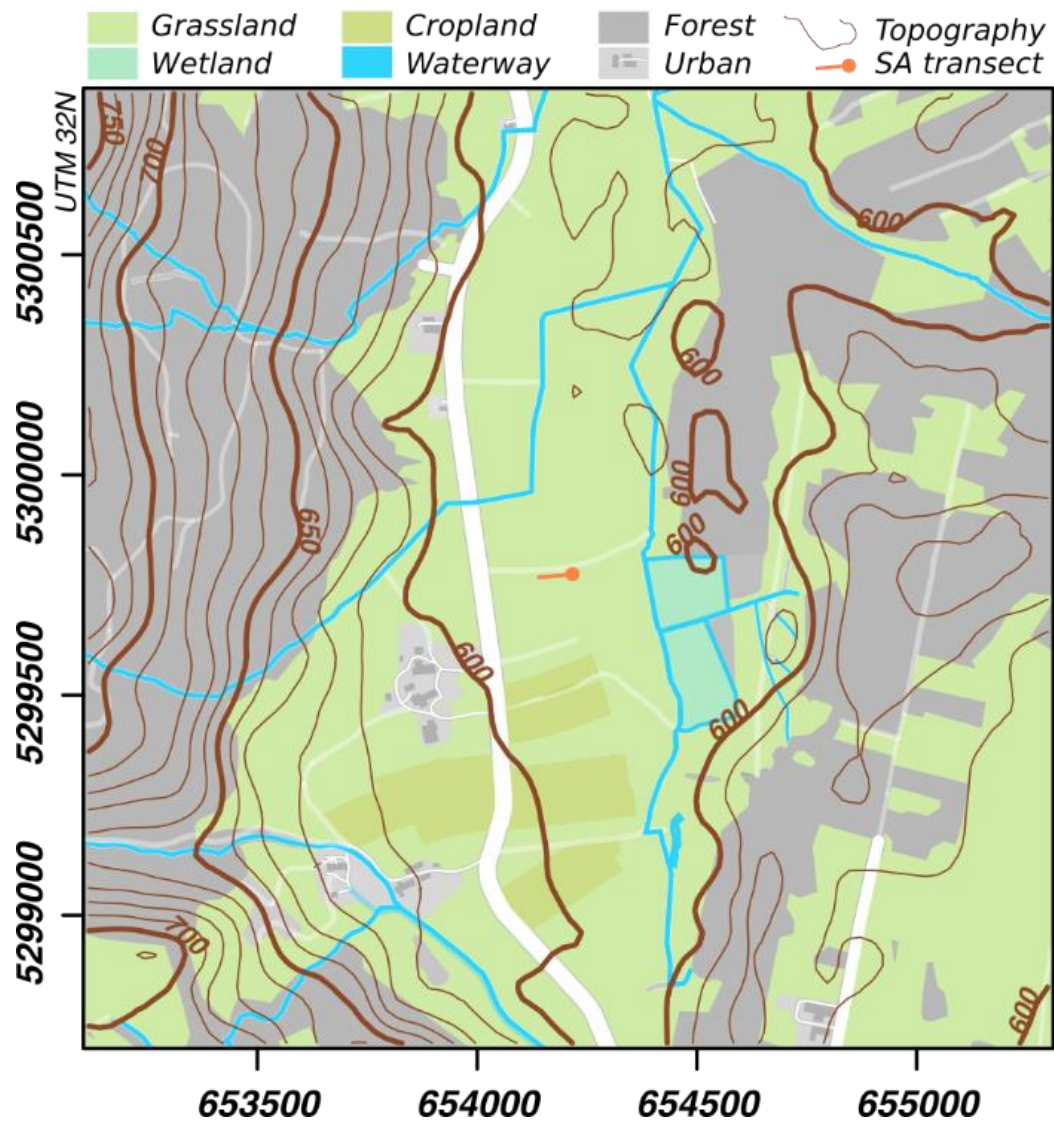
Figure 1: Location of the sonic anemometer (SA) transect at the DE-Fendt field. Map modified from Fig 1 in Zeeman et al. (2017)

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.

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,

10 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
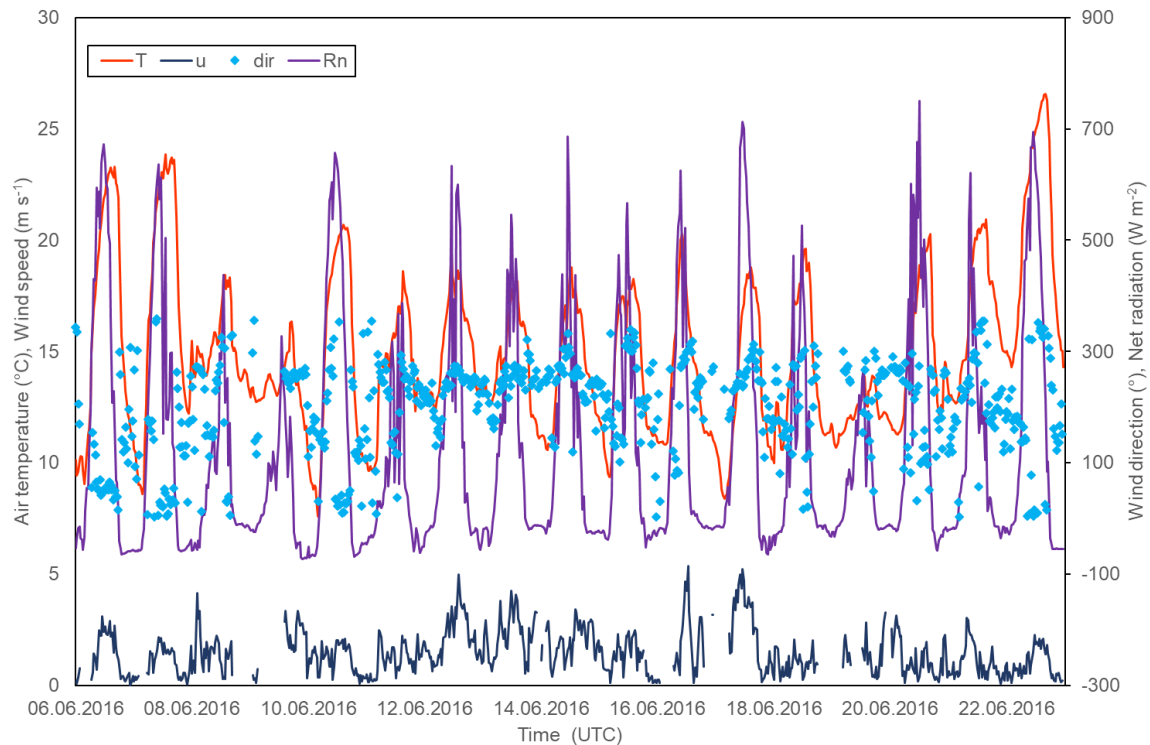
Figure 3: Meteorological elements during the intercomparison experiment; 30-min averages of air temperature (*T*) and net radiation (*Rn*) measured at 2 m, and wind speed (*u*) and direction (*dir*) from the DE-Fen site measured at 3.25 m.

2.2 Data processing

All data were processed 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 ($70° < \varphi < 110°$; $250° < \varphi < 290°$) 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 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
5 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
10 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
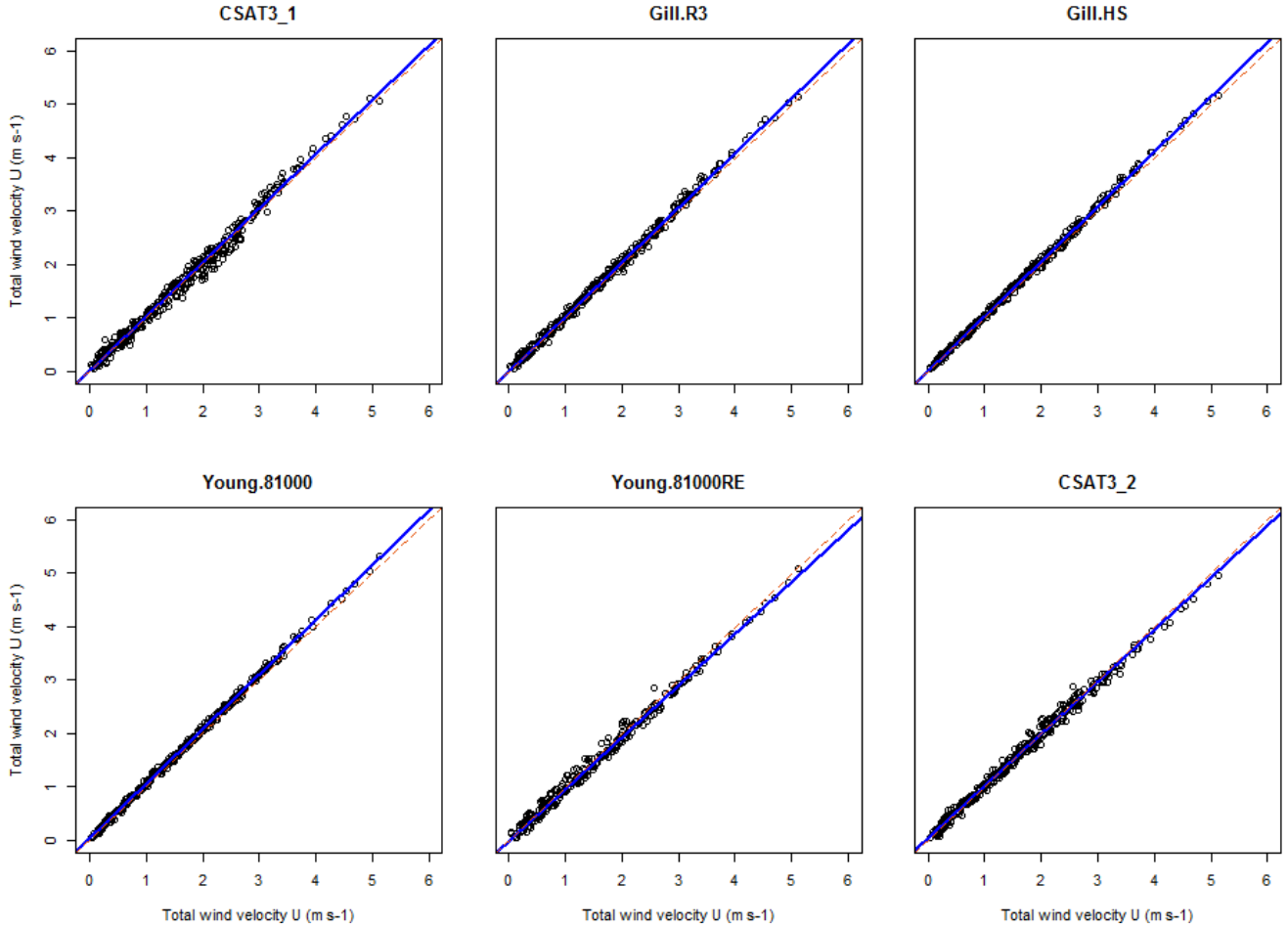
Figure 4: Comparison of the 30 min averaged total wind velocity measurements (etalon = METEK.uSonic.omni).

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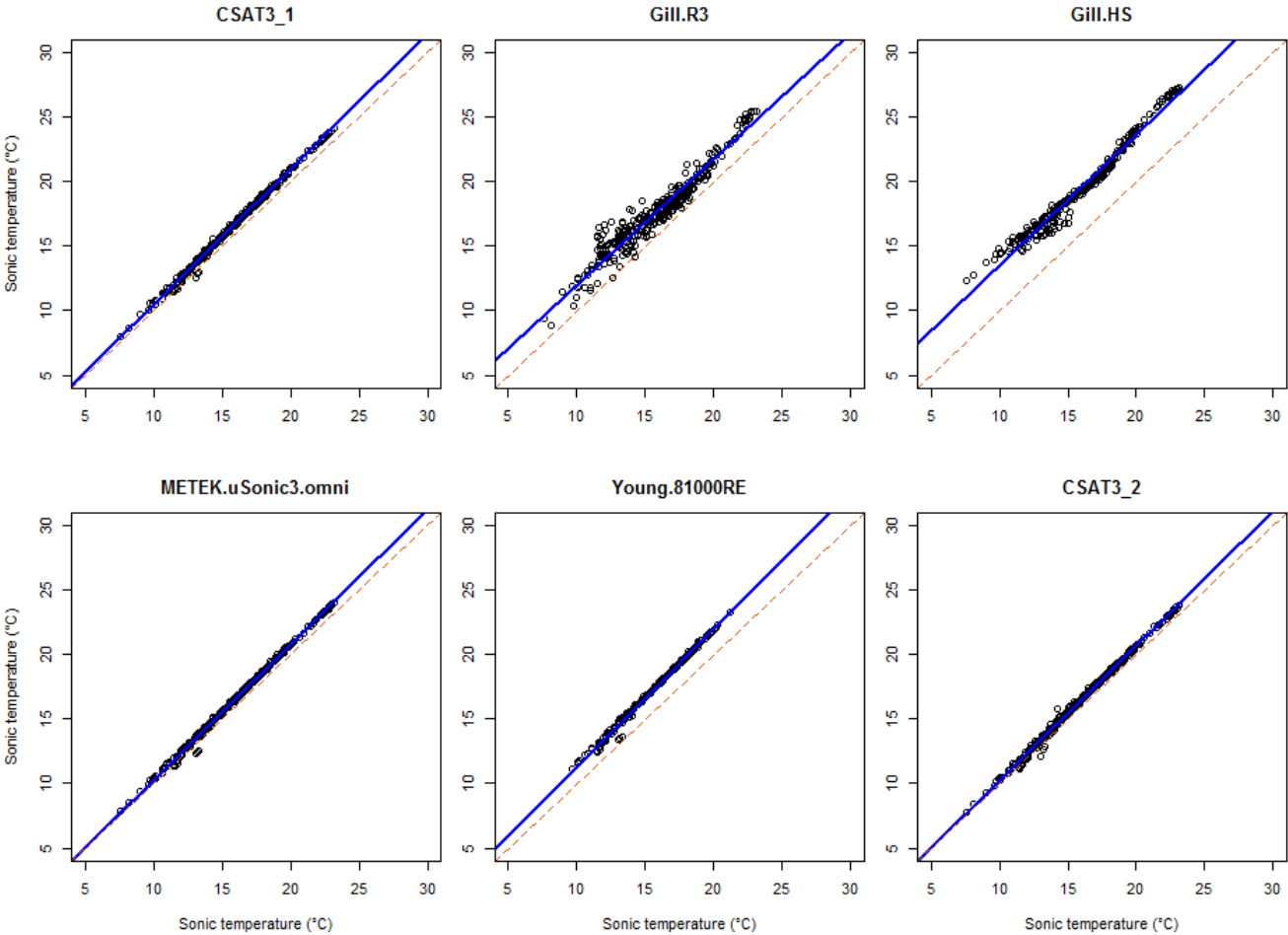
Figure 5: Comparison of the averaged sonic temperature measurements (etalon = Young.81000).

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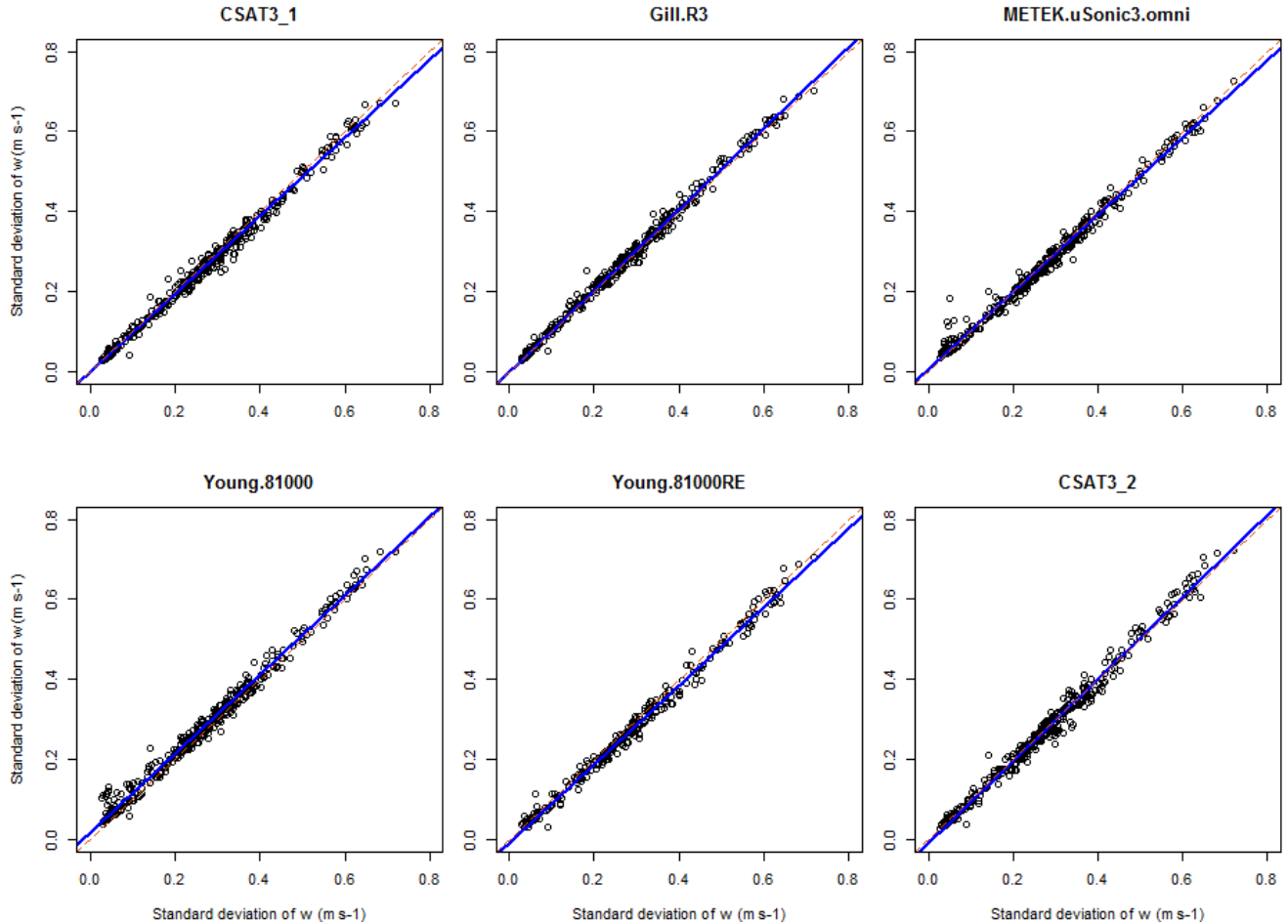
Figure 6: Comparison of the standard deviation of the vertical velocity component σ_w (etalon = Gill.HS)

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 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 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	1.06	1.05	1.04
bias (K)	0.00	0.00	-0.01	0.01	0.04	0.01
RMSE (K)	0.05	0.08	0.06	0.05	0.09	0.03
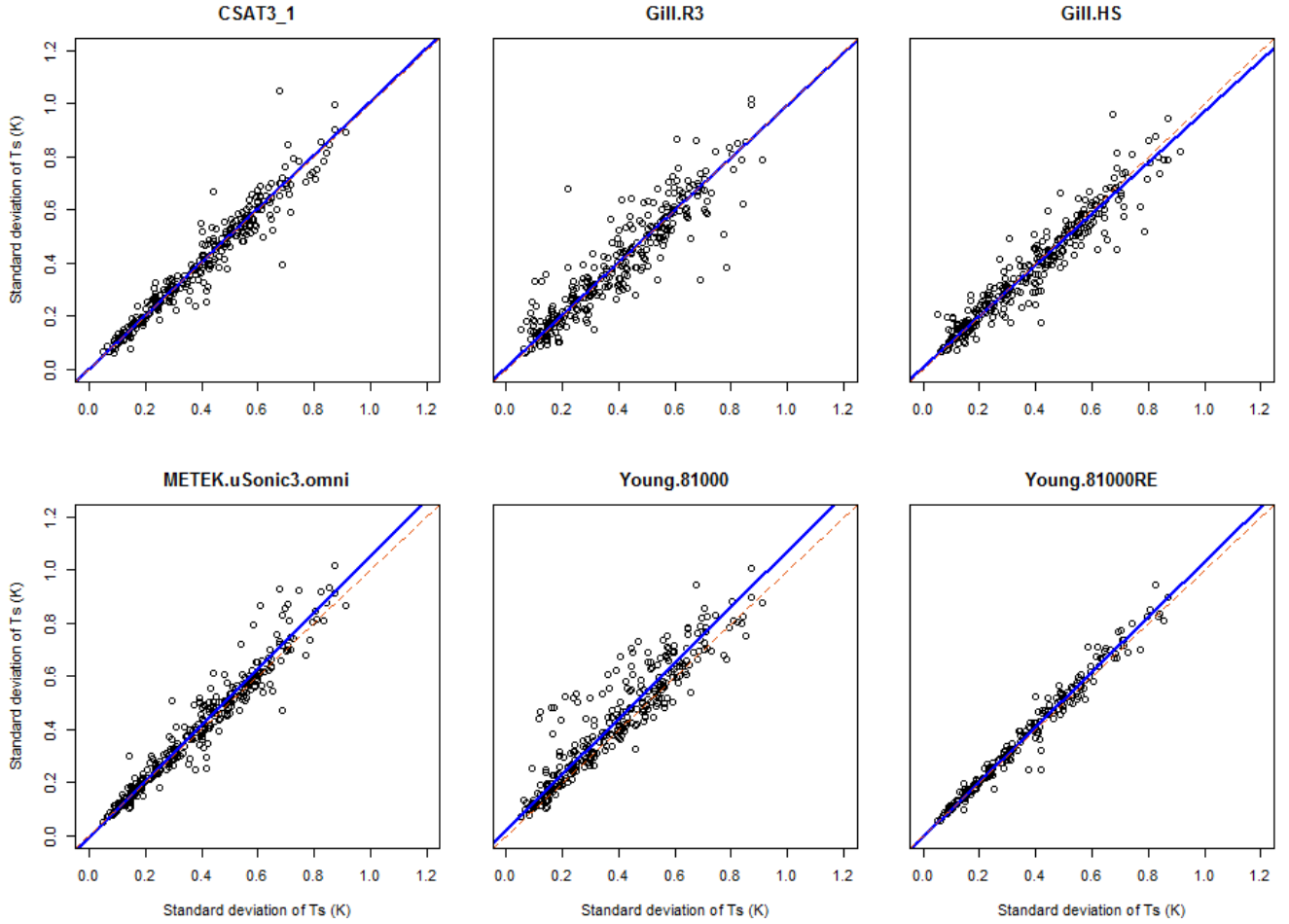
Figure 7: Comparison of the standard deviation of the sonic temperature σ_{Ts} (etalon = CSAT3_2).

3.5 Friction velocity

Friction velocities ranged between 0 m s^{-1} and almost 0.6 m s^{-1} during our experiment. Although the Young.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.
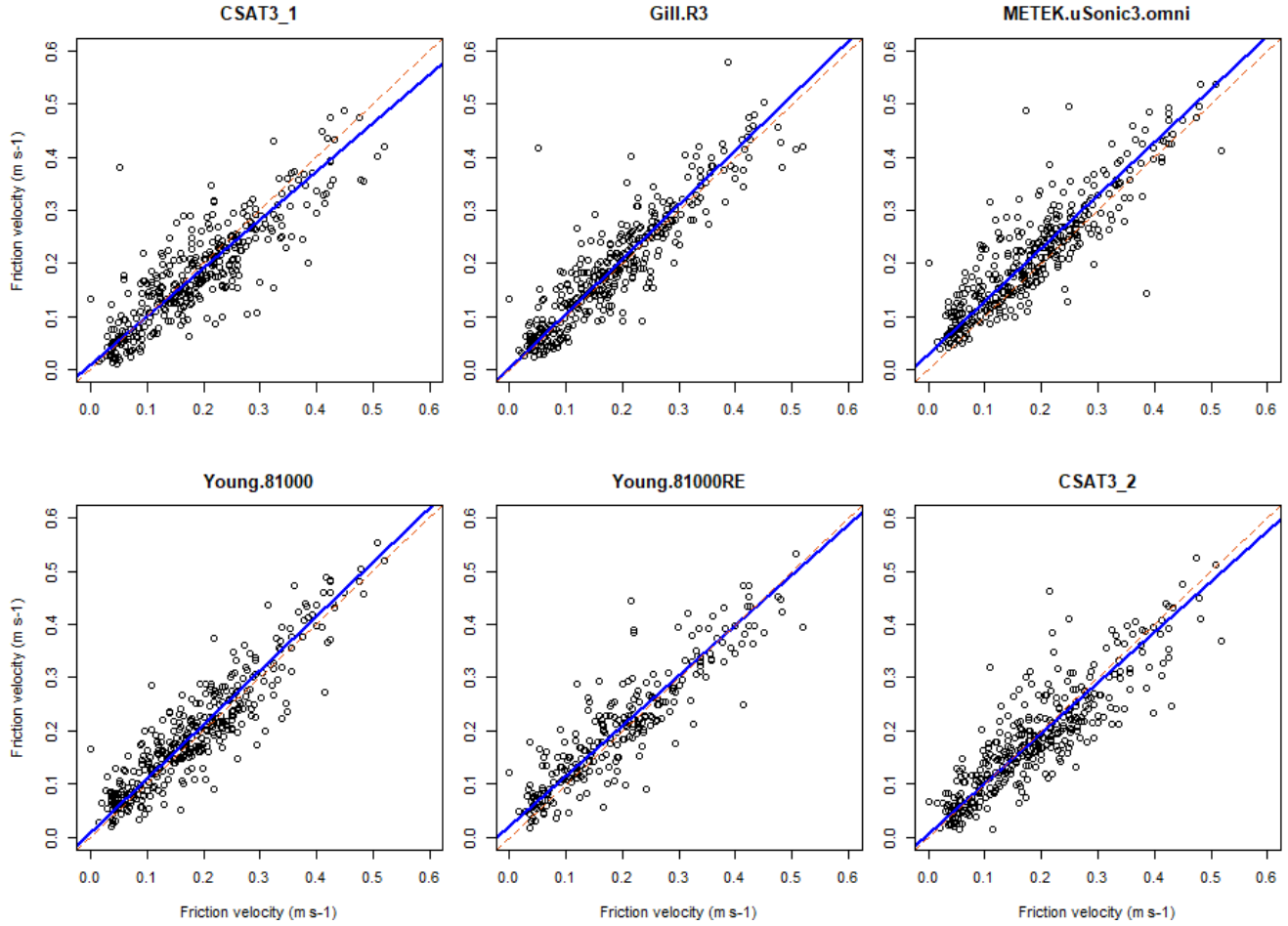
Figure 8: Comparison of the friction velocity measurements (etalon= Gill.HS).

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	0.91	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 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). 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.
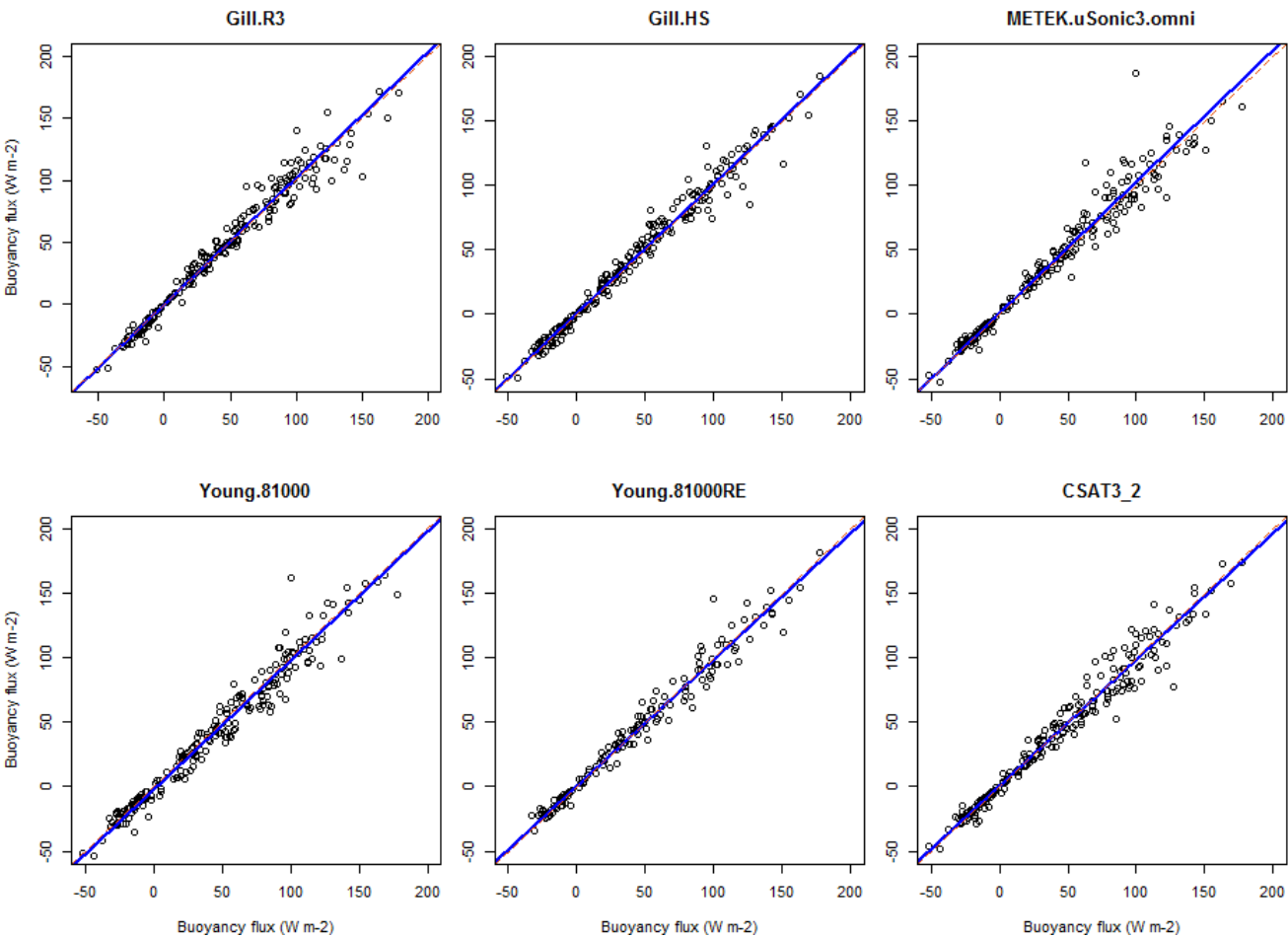
Figure 9: Comparison of buoyancy flux measurements (etalon = CSAT3_1)

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 systems collect independent samples of an homogeneous turbulence field (Richardson et al., 2012). The stochastic error 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 is as good as the agreement with other sonic anemometer models.

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. (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.

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.

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 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 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.

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). Hence, from our results we cannot derive a classification of the tested sonic anemometers in different quality levels,

10 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, 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

15 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

20 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 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

25 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

30 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

5   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

10   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,

15   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.

20   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.

25   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,

30   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.

5   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.

10 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

15 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,

20 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,

25 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.

30 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.

5   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.

10 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.

15 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,

20 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

25 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.,

30 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.

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.