Agricultural and Forest Meteorology 334 (2023) 109437

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

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

A study of the role of seven historically significant fast-response hygrometers and sensor calibration on eddy covariance H_2O fluxes and surface energy balance closure☆

John M. Frank [a,*], William J. Massman [a]

[a] USDA Forest Service, Rocky Mountain Research Station, 240 W. Prospect Rd., Fort Collins, CO, 80526

ARTICLE INFO

Keywords:
Evapotranspiration
Surface energy balance
Hygrometer
Eddy covariance
Ecosystem fluxes

ABSTRACT

Water vapor fluxes are a major component of the surface energy balance, hydrological cycle, and ecosystem processes. As newer sensors are developed for eddy covariance measurements, the role of different technologies and calibration uncertainties on H_2O fluxes and surface energy balance closure should be assessed. Here, a field intercomparison is described testing historically significant fast-response hygrometers: Lyman-α, KH2O, LI-7000, LI-7500, LI-7200, EC155, and EC150. This study occurred in southeastern Wyoming in September 2015. Three methods of calibration are investigated: standard calibration with a dew point generator and zero-air gas, referencing to a first principles slow-response hygrometer, and piecewise calibration between fast-response hygrometers based on blocks of time with consistent relationships between 5-minute data. Differences in fluxes among the infrared gas analyzers were largest when using the standard calibration (maximum difference of 15.9%, standard deviations of differences between 4.1–6.0%) with differences reduced by 0.3–0.6 × when the calibrations were referenced to a first principles hygrometer. The smallest differences occurred using the piecewise calibration, which emphasizes the intrinsic differences among the sensor technologies, where differences were further reduced by 0.04–0.09 × (maximum difference of 8.2%, standard deviations of differences between 2.3–3.0%). The Lyman-α and KH2O experienced severe drifts, though their piecewise calibrations were generally within the range of the others. The two most influential factors beyond calibration uncertainty were the oxygen correction for the ultraviolet absorption sensors, especially the KH2O, and the tube attenuation spectral correction for closed-path analyzers, mostly with the lower flow EC155. The effect of hygrometers on the energy balance closure was minor with some uncertainty associated with the standard calibration, though this was less than the change in closure associated with the choice of sonic anemometer. These results focus attention on the need to further develop accurate calibration methodologies that are suitable for all flux tower sites.

1. Introduction

Energy balance is one of a few fundamental laws in ecology (Odum, 1957), yet the inability of a vast majority of eddy covariance flux sites to adequately close their surface energy balance has been a persistent problem for the micrometeorological, ecological, and hydrological communities (Leuning et al., 2012; Mauder et al., 2020; Stoy et al., 2013; Wilson et al., 2002). Many causes of this imbalance have been investigated, ranging from advective transport (Morrison et al., 2022), energy storage (Reed et al., 2018), sensor biases (Frank et al., 2016b), etc. A unique subset of closure problems occurs in ecosystems where the imbalance varies in response to changes in vegetation and land cover. For example, seasonal trends in energy balance closure can occur in croplands as baren soils become covered with maturing vegetation that is ultimately harvested (Eshonkulov et al., 2019; Irmak et al., 2014, Ryken et al., 2022). Energy balance closure can also change following landscape altering disturbance, such as in forests after fire, logging (Amiro, 2001; Ha et al., 2015), and hurricanes (Barr et al., 2012). These situations highlight how the ability to measure each of the energy balance components can result in varying amounts of closure when an ecosystem enhances or diminishes different processes; this can be at odds with a common approach of closing the energy balance by

* Corresponding author.
E-mail address: john.frank@usda.gov (J.M. Frank).

https://doi.org/10.1016/j.agrformet.2023.109437
Received 27 May 2022; Received in revised form 8 February 2023; Accepted 20 March 2023
0168-1923/Published by Elsevier B.V.

maintaining the Bowen ratio and equally boosting sensible and latent heat fluxes (Twine et al., 2000). Rather, some studies have demonstrated that sensible and latent heat fluxes should be treated differently to achieve closure (Charuchittipan et al., 2014; Frank et al., 2019; Moorhead et al., 2019). If the turbulent fluxes contribute differently to uncertainties in the energy balance, then some of the imbalance could be attributed to the water vapor concentrations used in eddy covariance measurements.

The accuracy of fast response hygrometers and their effect on turbulent water vapor fluxes has been a focus of research along two themes: the differences between closed-path versus open-path sensors and generational improvements, i.e., the development of infrared gas analyzers (IRGAs), enclosed-path sensors, and the incorporation of the IRGA into the sonic anemometer structure. Though initial studies into the differences between open- and closed-path hygrometers simultaneously dealt with other differences in measurement technologies, i.e., ultraviolet vs. infrared absorption (Judd et al., 1993; Lee et al., 1994), most comparisons have focused solely on IRGAs (Anthoni et al., 2002; Burba et al., 2010; Haslwanter et al., 2009; Kosugi et al., 2007). Tube attenuation in closed-path sensors is often the major issue affecting accuracy between hygrometers (Lee et al., 1994; Fratini et al., 2012) along with the compounding effects of relative humidity (Ibrom et al., 2007), tube age (Mammarella et al., 2009), temperature (Runkle et al., 2012), and time lag (Wu et al., 2015). Recently, Zhang et al. (2021) focused on the relative humidity effect on all IRGAs. Generational improvements have produced hygrometers that are more stable, reliable, and with less noise and signal attenuation (Sonntag et al., 2021). Early generation fast-response hygrometers were based on ultraviolet absorption (van Dijk et al., 2003). With the advent of the modern IRGA, many studies have focused on the uncertainty in H_2O fluxes associated with sensors using different types of absorption (Judd et al., 1993; Lee et al., 1994; Martínez-Cob and Suvočarev, 2015; Mauder et al., 2006). The continued development of IRGAs has led to research on improvements in power consumption, instrument self-heating (Burba et al., 2008), optical stability, enclosed-path analyzers, etc., and ultimately the impact on H_2O fluxes (Burba et al., 2010; Kutikoff et al., 2021; Novick et al., 2013; Polonik et al., 2019).

These studies provide various metrics to judge the effects of sensor design on water vapor fluxes and have improved eddy covariance measurements, notably regarding the role of tube attenuation in the underestimation and uncertainty in H_2O fluxes and latent heat measurements. Yet, the role of calibration is typically not a focus in these studies. While manufacturers typically provide calibration instructions for their analyzers, within the micrometeorological community there is less of an accepted standard or procedure for calibrating fast-response sensors for water vapor than there is for carbon, e.g., network sponsored distribution of CO_2 calibration tanks (https://ameriflux.lbl.gov/tech/support-services/). Most studies mentioned above do not adequately specify the calibration routine of their fast-response hygrometers even when their CO_2 calibration, which takes place on the same sensor, is described in detail. This leads to a general confusion about the uncertainty in water fluxes; how much can be attributed to sensor technologies and how much is related to calibration? Fratini et al. (2014) developed a theoretical approach to correct errors arising from sensor drifts and demonstrated that sizeable differences in H_2O fluxes can be mitigated when uncertainties in calibration are addressed. A broadening of this concept across the spectrum of fast-response hygrometers would benefit the micrometeorology, ecological, and hydrological communities in studies of water vapor fluxes (evapotranspiration and sublimation), land surface interactions, and the surface energy balance.

To better understand the role of calibration uncertainty and hygrometer technologies on eddy covariance H_2O fluxes and surface energy balance closure, we investigate the following research questions. To what extent do standard calibration routines of fast-response hygrometers cause uncertainty in water vapor fluxes or energy balance closure? Are there fundamental differences among historically significant closed-path and open-path sensors that result in substantial uncertainty in H_2O fluxes or energy balance? To address these, we conducted a multiweek field study at a forested flux tower site during the end of the summer growing season where we deployed seven historically significant fast-response hygrometers. We investigate standard calibration procedures and introduce a novel piecewise calibration approach that uses similarity of data (quantiles and medians) over blocks of time to minimize the uncertainty in calibration. We evaluate differences in eddy covariance H_2O fluxes and energy balance closure due to these different hygrometers and their calibrations.

2. Methods

2.1. Sensors and study area

We investigate seven historically significant fast-response hygrometers from roughly three eras featuring both open- and closed-path technologies. The first-generation ultraviolet absorption hygrometers include a closed-path Lyman-α (HHeLM-L, Resonance Ltd., Barrie, ON, Canada) described by Beaton and Spowart (2012) and an open-path krypton hygrometer (KH2O, Campbell Scientific, Inc., Logan, UT, USA), the second-generation infrared gas analyzers include a closed-path LI-7000 (LI-COR, Inc., Lincoln, NE, USA) and an open-path LI-7500 (LI-COR, Inc.), and the third-generation infrared gas analyzers include a closed-path LI-7200RS (henceforth LI-7200, LI-COR, Inc.), a closed-path EC155 (i.e., the hygrometer portion of the CPEC, Campbell Scientific, Inc.) and an open-path EC150 (i.e., IRAGSON, but not integrated with a sonic anemometer, Campbell Scientific, Inc.). Eddy covariance fluxes (i.e., H_2O/latent heat and sensible heat) were measured with these hygrometers paired with two sonic anemometers (3Vx-probe, Applied Technologies, Inc., henceforth ATI, Longmont, CO, USA and CSAT3A, henceforth CSAT3, Campbell Scientific, Inc.). This study also utilized two slow response hygrometers, a cavity ring-down spectroscopy water vapor isotope analyzer (WVIA 912–1004–0000, Los Gatos Research, Inc., henceforth LGR, San Jose, CA, USA) and a capacitance humidity sensor (083D inside 076B-4 radiation shield, Met-One Instruments, Inc., Grants Pass, OR, USA).

The study occurred from 11 September to 2 October 2015 and was located atop of the glacier lakes ecosystem experiments site (GLEES) AmeriFlux scaffold (US-GLE) in the mountains of southeastern Wyoming, USA (Frank et al., 2014; Frank and Massman, 2021; Musselman, 1994). This is a high-elevation spruce-fir forest with substantial tree mortality from a spruce beetle outbreak resulting in a median canopy height of 7 m reaching a maximum of 18 m (Burns et al., 2021). The Lyman-α, KH2O, LI-7000, and LI-7200 were installed on 11 September while the EC155, EC150, and CSAT3 were installed on 26 September 2015; all were subsequently removed on 2 October 2015. Each of these sensors or their inlet tubes were mounted at 25.2 m on the scaffold, which was above all floors, walls, and guardrails of the structure (Fig. 1). The sensors were pointed due west, with the CSAT3 and KH2O furthest west and the others between 0.15 to 0.28 m east; the prevailing wind during the study was from 268° with 80% of the data between 215° and 314° All closed-path sensors were mounted on top of the scaffold in weatherproof enclosures, thus allowing for short tube lengths, i.e., 2 m of 1.6 mm inner diameter (ID) Bev-A-Line IV (Thermoplastic Processes, Inc., Georgetown, DE, USA) for the Lyman-α and LI-7000, 0.48 m of 1.9 mm ID stainless steel tubing for the LI-7200, and a 0.65 m long and 1.1 mm ID intake for the EC155, all unheated. The LI-7500, ATI 3Vx-probe, and LGR were part of the long-term GLEES AmeriFlux instrumentation with these sensors or their inlet tubes mounted at 22.65 m on the scaffold (Fig. 1). The Lyman-α, KH2O, LI-7000, and LI-7200 were measured at 20 Hz on a CR3000 datalogger (Campbell Scientific, Inc.). The LI-7000 was recorded via its analog output until 22 September when the digital communications were used; this resulted in minor differences in H_2O values and time lag as recorded by the datalogger. The EC155,

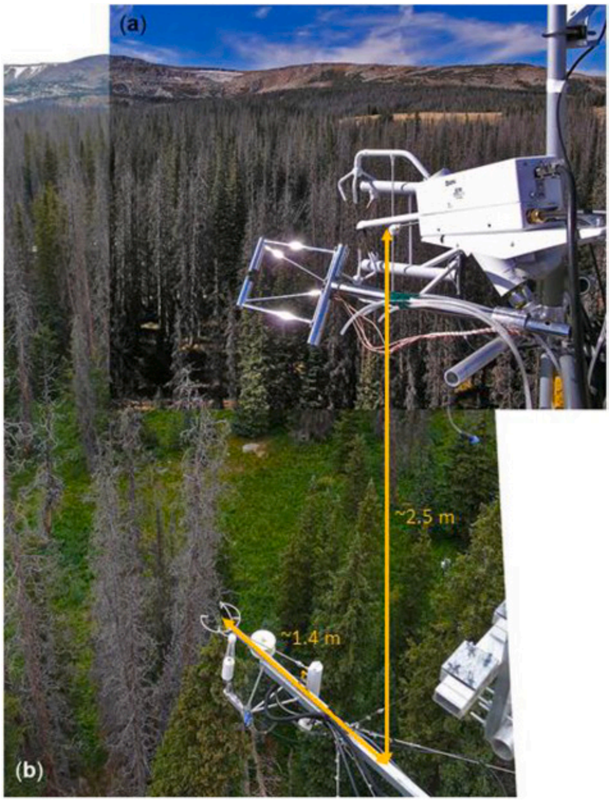

Fig. 1. (a) The hygrometer instrumentation on 26 September 2015 after installation of the EC150, EC155, and CSAT3. From 11 to 26 September only the LI-7200, KH2O, along with sample tubing for the Lyman-α and LI-7000 were mounted above. (b) A photo of the main US-GLE eddy covariance system taken from the same vantage on 16 September 2016 showing the LI-7500 and ATI 3Vx-probe sonic anemometer mounted 22.65 m above the ground. In 2015, the hygrometer instrumentation mounted ~ 2.5 m above and ~ 1.4 m behind this equipment.

EC150, and CSAT3 were measured at 10 Hz on a separate CR3000 datalogger. The LI-7500 and ATI 3Vx-probe were measured at 20 Hz on a third CR3000 datalogger as part of the US-GLE eddy covariance system. The two 20 Hz CR3000 dataloggers were synchronized every 6 h; the US-GLE datalogger served as the reference. The 10 Hz CR3000 datalogger was synchronized when it was initially deployed on 26 September 2015. All other supporting meteorological data were measured recorded every 10-seconds as part of the US-GLE AmeriFlux site and are described in Frank et al. (2014).

Data from each hygrometer were screened for anomalous values by comparing 10-s water vapor concentrations (i.e., mole fraction or molar mixing ratio, χ_v) to the Met-One capacitance humidity sensor. After various calibrations were applied to the hygrometers (see the following sections), 5-minute eddy covariance H_2O fluxes were calculated as described by Frank et al. (2014). Briefly, this includes calculations of covariances, planar fit rotation, time lag adjustments, spectral corrections for block averaging, sensor path averaging, horizontal sensor separation, and tube attenuation (Massman, 2000), IRGA self-heating corrections for the LI-7500 (Burba et al., 2008; Frank and Massman, 2020), and Webb-Pearman-Leuning corrections (Webb et al., 1980) using inhouse code in MATLAB (MATLAB, 2018). The spectral corrections for tube attenuation were conducted using attenuation coefficient approach of Massman (1991) where the Reynold's number is determined by tube dimensions, flow rate, and the kinematic viscosity of air (Massman, 1999). Above the critical Reynolds number of 2300, e.g., when flow becomes turbulent, the attenuation coefficient was estimated using the analytical approximation of Massman and Ibrom (2008); below this threshold, when flow is laminar, the spectral correction time constant was determined from the half-power attenuation frequency (Luening and King, 1992). We estimated spectral time constants for flow through both the tubing and the cells of the closed-path hygrometers. For covariances with a mismatch in sampling rate between the sonic anemometer and hygrometer, e.g., 10 Hz versus 20 Hz, etc., the sonic anemometer was resampled to match the sampling rate of the hygrometer. For the CSAT3 sonic anemometer, the transducer shadowing algorithm of Wyngaard and Zhang (1985) was applied as described in Horst et al. (2015). Flux processing was done using the various hygrometer calibrations, with and without the O_2 correction for the Lyman-α and KH2O hygrometers (van Dijk et al., 2003), and with and without the tube attenuation correction for closed-path hygrometers. For analysis of the effect of vertical separation between the sonic anemometer and scalar sensors on flux calculations, 5-minute sensible heat fluxes (H) were calculated for all combinations of vertical wind velocity and sonic virtual temperature between the two anemometers. For the energy balance analysis, 5-minute sensible heat fluxes were calculated as described by Frank et al. (2014) for both sonic anemometers, with the 5-minute available energy calculated using the net radiation (R_n) and soil heat flux (G) from the US-GLE AmeriFlux site and the energy storage in the canopy (S) calculated as described by Frank et al. (2019). R_n was measured with four-component radiometers (PSP and PIR, Eppley Laboratory, Newport, RI, USA), G was measured with two replicate soil heat-flux plates 1 m apart at 0.09 m depth (HFT-1, Radiation Energy Balance Systems, Inc.) and corrected to surface heat flux (i.e., 0.00 m depth) via the gradient of soil temperature at 0.03 and 0.09 mm measured near each plate (24 AWG type-T thermocouple wire insulated with Omegabond 101, Omega Engineering, Inc.) and multiplied by the specific heat of soil, which was in turn estimated from site specific soil mineral properties and the average gradient of soil water content at 0.05 and 0.10 m depth at seven replicates profiles within 6 m of the plates (5 with CS605/TDR100 and two with CS616 probes, Campbell Scientific, Inc.). The rate change of energy storage included the sensible and latent heat storage fluxes into the canopy airspace and the heat storage fluxes into the canopy leaves and trunks based on the multilayer cylindrical canopy methodology of Haverd et al. (2007) and adapted for GLEES using site-specific plant and biometric measurements.

2.2. Hygrometer calibrations

As described below, all hygrometers were deployed with a “raw” or off the shelf calibration, which typically was set in the laboratory. During the study we conducted standard calibrations in-situ on the scaffold using a dew point generator and zero air gas. In post-processing we conducted a comparison against a first principles sensor that was deployed at the same time of the experiment. Finally, in post-processing we conducted a piecewise calibration that identified periods of stability between hygrometers to construct a series of calibration coefficients.

2.2.1. Standard calibration

All fast-response hygrometers were calibrated on the scaffold with a dew point generator (LI-610, LI-COR, Inc.) and tank of ultrapure air (henceforth zero-air), i.e., standard calibration, at least once during the study. Water vapor concentrations were generated corresponding to dew points between 0 °C and the ambient temperature, i.e., generally at 4 or 5 different concentrations, with each setting allowed to flow for ~ 2 min. The flow rate on the LI-610 was set around 2 liters per minute (LPM), which was allowed to flow freely through the calibration chambers of the open-path sensors. The internal pressure of the LI-610 was estimated to be the ambient pressure from the US-GLE AmeriFlux site. For closed-path sensor calibrations, the LI-610 restricted the normal flow through the sensors (i.e., 12 LPM for the Lyman-α and LI-7000, 18.5 LPM for the LI-7200, and 7 LPM for the EC155), causing the LI-7000 and LI-7200 cell pressures to drop by ~ 2 kPa during calibration. For the LI-7200, calibration was also attempted with the sensor pump bypassed, causing the head pressure to increase by ~ 1 kPa; the difference in calibration slope caused by under-pressuring vs. over-pressuring the sensor head was ~ −4%. For the EC155, the calibration was done with the pump bypassed, causing the head pressure to increase by ~ 1 kPa. The standard calibration equations were obtained by using linear regression between the water vapor concentration of each hygrometer relative to the water vapor concentration generated by the LI-610/zero-air tank using the MATLAB curve fitting tool (MATLAB, 2018). We used the reduced major axis (RMA), e.g., the geometric mean of the slope from the regression with the LI-610 as the independent variable with the inverse of the slope from the regression with the other hygrometer as the independent variable, to estimate calibration coefficients that are symmetric (i.e., relatively the same when comparing measurement 1 to 2 or vice versa) while considering error in both measurements (Smith, 2009). The zero-air point often had a disproportionate effect on the standard calibration equations; thus, regressions were conducted with and without the zero-air, with the ramifications of this interpreted in the discussion. Because of frequent drifts during preliminary tests in the Lyman-α calibration, this hygrometer was equipped with a dedicated LI-610 that performed automated multipoint calibrations throughout the experiment; ultimately this was insufficient and none of the Lyman-α standard calibration data were used. Though the KH2O was calibrated with a LI-610 and its windows were cleaned during each site visit, it also experienced rapid sensor drift, and similarly, none of the standard calibration data were used.

2.2.2. Comparison against the first principles based LGR

After applying the standard calibration, each of the 2nd and 3rd generation hygrometers were compared to the LGR, i.e., a first principles sensor derived from fundamental laws of physics and that does not require calibration, with the goal to evaluate the linearity of each sensor and its stability over time. Because the LGR is a slow-response sensor, the data for each fast-response sensor were processed with a 90 s time-constant exponential moving average to mimic the slow response of the LGR. A time-constant was initially chosen to correspond to the flow rate and cell volume of the LGR; this was iteratively adjusted to a longer time to improve the correlation between the hygrometers. The LGR was also configured to periodically sample a vertical profile on the GLEES scaffold, which resulted in sporadic data from the top of the scaffold that occurred in blocks of 5–7 min every half hour. This also led to numerous occasions where the sample from the top of the scaffold appeared contaminated with the sample from the other inlets, i.e., during high humidity moisture was difficult to purge from the measurement cell. Thus, the LGR dataset was screened to include only the subset of ≤ 7-minute blocks when the LGR clearly was measuring the same signal as the other hygrometers. The median low-pass filtered concentrations measured from each hygrometer during each of these time blocks were used for the comparison. The water vapor concentration of the LGR was predicted as a function of the low-pass filtered water vapor concentration from each fast-response hygrometer. For analyses of water vapor fluxes, we defined the calibration with respect the LGR as the first order polynomial fit, i.e., Eq. (1) with $N = 1$, $M = 0$, $b_0 = 0$. As a diagnostic to quantify the stability of each fast-response hygrometer over time, we considered higher order polynomials of water vapor concentration in addition to time (t) such that

$$\chi_{v,LGR} = \sum_{n=0}^{N} c_n \chi^n_{v,hygrometer} + \sum_{m=0}^{M} b_m t^m \qquad (1)$$

where the values c_n and b_m are regression coefficients and the degrees of the polynomials, N and M, were selected by minimizing the root-mean-square-error (RMSE). These regressions were performed using the MATLAB curve fitting tool (MATLAB, 2018).

2.2.3. Piecewise calibration

For the piecewise calibration, the fast-response hygrometers are compared to a reference, and for each uninterrupted block of time when the data between the hygrometer and the reference have a consistent response, a linear calibration between them is applied. In general, we used the standard calibrated LI-7200 as the reference. The following algorithm achieves a piecewise calibration between two times series of water vapor data using molar mixing ratio or mass density. (1) For all potential block lengths and start/end times to be considered, perform linear regression between the subsets of water vapor from the two sensors; create a matrix of regression results (e.g., R^2, model error variance, range of data, etc.) arranged by block length and start time. For these regressions we defined the reference, i.e., LI-7200, as the dependent variable and calculated the range of its data. (2) Define a test metric to determine if the quality of fit for each proposed block. While this could be the R^2 or root-mean-square-error (RMSE) of the regression, we defined this as the model error variance divided by the range of data. Using this metric, smaller values correspond to tighter fits to a regression line and/or blocks that span larger ranges of ambient background humidity. This values of this metric are mmol mol^{-1} or g m^{-3}, whereas a similar approach using the model error standard deviation divided by range of data would produce a unitless metric. (3) Select a threshold to evaluate the test metric. We evaluated progressively smaller thresholds (i.e., 0.004, 0.002, 0.001, 0.0005 mmol mol^{-1}, etc.) and selected an optimal value based on the overall performance of the piecewise calibration (see step 7 below). (4) Search the matrix of regression results for blocks where the test metric falls under the threshold, beginning with the longest blocks and progressing towards smaller blocks. For every block that is selected as meeting the search criteria, every smaller block that includes any of the timesteps are removed from further consideration. The search continues until a minimum block length is considered and no more blocks are found. Data not assigned to a block typically occur when there is noticeable sensor drift. (5) Create a timeseries of piecewise calibration coefficients by performing regression for each selected block between the water vapor data from the two sensors. We used the RMA for these regressions, with the exception of the Lyman-α which we fit with a 2nd order polynomial, and hence was not adaptable to the RMA. As the timeseries of calibration coefficients will have gaps corresponding to data that were not assigned to a block, one variation is to linearly interpolate the coefficients for all timesteps. While we discuss implications of doing this, we do not use interpolated results in our statistical analyses. For consistency when comparing results between calibration methods, we removed these non-assigned data points from the other datasets (i.e., standard calibration, adjusted to the LGR, etc.).

(6) Apply the piecewise calibration coefficients to the water vapor to generate a calibrated dataset. (7) Evaluate the overall performance of the piecewise calibration, using diagnostics such as the number of blocks identified, the percent of data that was incorporated into a block, and R^2 or RMSE when comparing the piecewise calibrated hygrometer data to the reference. These diagnostics can be compared for various thresholds (step 3) to determine an optimal application, as discussed in section 4.4. The focus of our paper was to apply the piecewise calibration using a dataset of 5-minute quantiles of water vapor data (5, 10, 25, 50, 75, 90, and 95%) which can characterize the turbulent responses of the fast-response hygrometers. To evaluate the feasibility of applying this algorithm with slow-response hygrometers we repeated the procedure on a dataset of 5-minute medians. All work was done in MATLAB (MATLAB, 2018) with an example of the source code presented in the supplementary material Text S1. For open-path sensors with a native measurement of vapor density, the piecewise calibration coefficients (step 5) are still determined using RMA from the corresponding molar mixing ratio, based on the assumption that all hygrometers should measure the same mole fraction; after the RMA is performed, the piecewise calibration coefficients are transformed to correspond to vapor densities so they could be applied to the original hygrometer measurements. By assigning the standard calibrated LI-7200 as the reference, the absolute accuracy of all piecewise calibrated hygrometers is equal to the absolute accuracy of the standard calibrated LI-7200.

Finally, during the last few days of the study, much of the LI-7200 data was lost due to power failures. Because of this, a secondary the piecewise calibration was conducted for the LI-7500 and EC-150 using the EC-155 as the reference, where additional calibration coefficients were added when the LI-7200 was missing. These secondary calibration coefficients were adjusted to account for the piecewise relationship between the EC-155 and LI-7200. Finally, as the start/end times of the piecewise blocks typically occur when there is noticeable sensor drift, we allowed visual inspection and removal of calculated water fluxes only for the 5-minute periods immediately at the edges of the piecewise blocks.

2.3. Statistical analysis of eddy covariance H_2O fluxes, vertical separation of sensors, and surface energy balance closure

Bayesian statistics were used to estimate the relative differences in water vapor fluxes and energy balance closure ratios when using the various hygrometers, as well as to quantify the effect of vertical separation between sonic anemometers/scalar sensors. To quantify differences in H_2O fluxes, the statistical model predicts each 5-minute eddy covariance flux (F_v) from the seven hygrometers as a percent difference (α_1) from the "baseline" water flux ($\widetilde{F}_v$) based on the relationships

$$F_{vh,i} \sim \text{Norm}\left(\theta_{h,i}, \ \sigma^2\right) \text{ for } h = 1, \ldots, 7 \text{ and } i = 1, \ldots, \ 5988 \quad (2)$$

$$\theta_{h,i} = \widetilde{F}_{vi} \times (1 + 0.01\alpha_{1h}) \quad (3)$$

$$\widetilde{F}_{vi} \sim \text{Unif}\left(1.1 \ \times min\left(F_{vh,i}\right), \ 1.1 \ \times max\left(F_{vh,i}\right)\right) \quad (4)$$

$$\alpha_{1h} \sim \text{Norm}(0, \ 1) \quad (5)$$

$$\sigma \sim \text{Gamma}\left(0.011 \ \times \left(max\left(F_{vh,i}\right) - min\left(F_{vh,i}\right)\right), \left(0.011 \ \times \left(max\left(F_{vh,i}\right) - min\left(F_{vh,i}\right)\right)\right)^2\right) \quad (6)$$

where Norm(*E, Var*) and Gamma(*E, Var*) are normal and gamma distributions with expected value, *E*, and variance, *Var*, and Unif(min, max) are uniform distributions defined from a minimum to maximum value. The values $\widetilde{F}_v$, α_1, and σ (the standard deviation of the prediction) are statistical process parameters assigned prior probability distributions; the Bayesian analysis combines these with the likelihood of observing the data to estimate posterior probability distributions. These prior distributions are considered generally non-informative. The "baseline" water flux is conceptually the average of the fluxes from the seven hygrometers for every 5-minute period. The percent differences are conceptually the slope of the regression between the flux from each hygrometer relative to the "baseline" over the entire study period. For easier interpretation, the slope is presented as the percent difference from the 1:1 line, such as a hygrometer is said to be a few percent higher/lower than the average. This procedure allows for multiple comparisons between seven simultaneous datasets that have inconsistent sample sizes, while not requiring any one hygrometer to be identified as the absolute standard (note, this is not to be confused with using the LI-7200 as the reference for the piecewise calibration). In this approach, the "baseline" will be influenced by changes in the data, i.e., testing different calibrations, rendering the α_1 parameters as somewhat arbitrary when considered individually. The differences among the various hygrometers between α_1 parameters describe whether a specific hygrometer measures a relatively smaller or larger water vapor flux, etc. Because the F_v data from each hygrometer is predicted with the same σ, each sensor contributes equally to the error in the Bayesian analysis. We minimize the risk of biasing the "baseline" with an erroneous sensor by pre-screening all data as described in Section 2.1. Otherwise, if a hygrometer measures fluxes that are reasonable except in their magnitude, this will influence the magnitude of the "baseline" but not the shape of its timeseries; the values of the individual α_1 parameters will also be influenced but not the relative differences between them. The analysis of vertical separation of anemometers/scalar sensors follows the same procedure, except the data analyzed are sensible heat flux and the "baseline" is conceptually the average of the four combinations of vertical wind velocity (w) and sonic virtual temperature (T_s) between the ATI 3Vx-probe and CSAT3 anemometers. The analysis of the energy balance closure ratio predicts the 5-minute turbulent fluxes from the available energy, i.e., there is no "baseline" in this analysis as there is a true independent variable to be used as the predictor variable. Also, to be consistent with much of the literature, the α_1 parameters quantifying the difference between hygrometers is the energy balance closure ratio with an expected value = 1, i.e., $\alpha_{1h} \sim \text{Norm}(1, \ 1)$.

For all Bayesian analyses we utilized Markov Chain Monte Carlo (MCMC) methods (Kruschke, 2014). For the energy balance analysis there are only eight process parameters (seven closure ratios, i.e., one for each hygrometer, one standard deviation of the prediction,). For the water vapor flux analyses there were 5996 process parameters (seven percent differences, i.e., one for each hygrometer, one standard deviation of the prediction, and 5988 "baseline" parameters for each 5-minute period) and for the vertical separation of anemometers/scalars there were 5994 parameters (i.e., four percent differences, i.e., one for each sensible heat flux combination, etc.). Regarding the 5988 5-minute periods, if data were missing from all sensors, then the posterior distribution was simply equal to the prior distribution. Finally, all Bayesian analyses were conducted using R (version 4.0.5), (R Core Team, 2015) within Rstudio (version 1.1.456), (Rstudio Team, 2015), using packages CODA (Plummer et al., 2006) and rjags (Plummer, 2022) with the MCMC analysis implemented by JAGS (version 4.0.0), (Plummer, 2003) using 25 MCMC chains, adapting each chain for 100 steps, running 400 steps for burn-in, and running another 1000 steps to estimate the posterior distributions. To reduce correlation between steps, the final MCMC chains were thinned to keep 1 of every 10 steps, i.e., a total of 100 steps per chain = 2500 saved steps, which were combined to estimate the posterior distributions. All MCMC chains were visually inspected using trace plots to confirm burn-in and convergence has been achieved. In the case of the energy balance analysis with the ATI 3Vx-probe, the burn-in was increased to 900 steps. We interpret parameters as significantly different when their 95% credible intervals do not overlap (Lu et al., 2012).

To examine the consequences of processing fluxes in 5-minute periods, we compared the sensible and latent heat fluxes from the ATI 3Vx-probe and the standard calibrated LI-7500 against H and F_v from the US-GLE dataset (Frank and Massman, 2021) which is sourced from the exact data and reported as 30-minute fluxes. For this comparison we averaged the 5-minute fluxes corresponding to each 30-minutes period and conducted a linear regression with the US-GLE data using the reduced major axis.

3. Results

3.1. Standard calibration versus dew point generator

A standard calibration was conducted on various dates using the dew

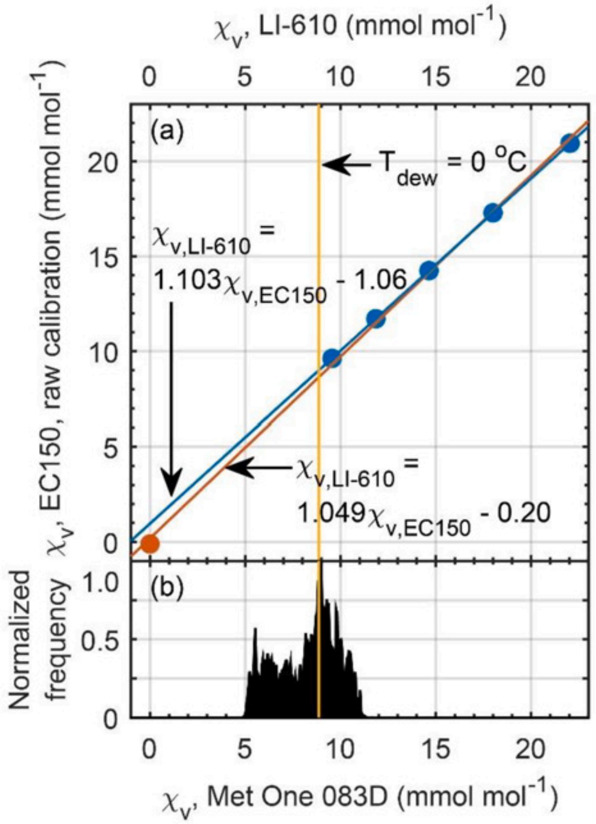

Fig. 2. The standard calibration of the EC150 hygrometer performed at GLEES on 26 September 2015. (a) The LI-610 generator was used to create five water vapor concentrations (χ_v) at dew points between the ambient temperature and freezing (i.e., 1, 4, 7, 10 and 13 °C, blue dots) and a zero-air tank was used to generate a sixth point (i.e., red dot). The regression slope is ~ 5% larger when the zero-air point was included, though the root-mean-square error increases by a factor of 58 ×. (b) The majority of vapor concentrations from the slow-response meteorological sensor measured from 26 September to 2 October 2015 (i.e., the timeframe when the EC150 was deployed) are drier than the lowest concentration achievable by the LI-610 generator.

point generator and zero-air tank. Fig. 2 shows for the EC150 how the values from the standard calibration result in different regression fits depending on the inclusion of the zero-air data. The resulting standard calibration equations for the 2nd and 3rd generation hygrometers are listed in Table 1. The effect of repeating the standard calibration a week later, regardless of the inclusion of the zero-air data, was to reduce the regression slopes for the LI-7000 and LI-7200 by −0.061 to −0.088; the LI-7500 was relatively unaffected with a maximum change of −0.009. The effect of including the zero-air data, regardless of the date of the calibration, was to reduce the slopes for the EC155 and EC150 by at least −0.049 and to change the magnitude of the slopes for the LI-7000 and LI-7500 by at least 0.017; the LI-7200 was less affected with a maximum change of 0.020. During the 17 September calibrations fewer points were generated due to the lower air temperature, this ultimately led to us omitting these calibrations from subsequent analyses. For calibrations from the other dates the RMSE of the regression fits without zero-air were 0.034, 0.033, 0.008, 0.006, and 0.005 mmol mol^{-1} for the LI-7000, LI-7500, LI-7200, EC155, and EC150, respectively; these increased to 0.117, 0.122, 0.032, 0.180, and 0.298 mmol mol^{-1} with the inclusion of the zero-air. Thus, for the EC155 and EC150, the inclusion of the zero-air data increased the RMSE of the regressions by a factor of 33 × and 58 ×, respectively; for the other hygrometers the increases in RMSE were between 3.5 × and 3.8 ×.

Table 1
Calibration equations for the 2nd and 3rd generation hygrometers with the standard calibration using a LI-610 dew point generator and a zero-air tank on the GLEES scaffold. Results for the Lyman-α and KH2O are not included because of severe instrument drift.

Hygrometer	Generation	Date of calibration	Without zero-air	With zero-air
LI-7000	2nd	17 September 2015	$\chi_{v,LI-610} = 0.982\chi_{v,LI-7000} + 0.32$	$\chi_{v,LI-610} = 1.009\chi_{v,LI-7000} + 0.03$
		24 September 2015	$\chi_{v,LI-610} = 0.908\chi_{v,LI-7000} + 0.92$	$\chi_{v,LI-610} = 0.925\chi_{v,LI-7000} + 0.63$
LI-7500	2nd	17 September 2015	$\chi_{v,LI-610} = 0.994\chi_{v,LI-7500} - 0.77$	$\chi_{v,LI-610} = 0.961\chi_{v,LI-7500} - 0.38$
		26 September 2015	$\chi_{v,LI-610} = 0.989\chi_{v,LI-7500} + 0.25$	$\chi_{v,LI-610} = 0.970\chi_{v,LI-7500} + 0.57$
LI-7200	3rd	17 September 2015	$\chi_{v,LI-610} = 1.008\chi_{v,LI-7200} - 0.58$	$\chi_{v,LI-610} = 0.988\chi_{v,LI-7200} - 0.34$
		24 September 2015	$\chi_{v,LI-610} = 0.920\chi_{v,LI-7200} + 0.11$	$\chi_{v,LI-610} = 0.927\chi_{v,LI-7200} - 0.01$
EC155	3rd	26 September 2015	$\chi_{v,LI-610} = 1.095\chi_{v,EC155} - 0.81$	$\chi_{v,LI-610} = 1.046\chi_{v,EC155} - 0.14$
EC150	3rd	26 September 2015	$\chi_{v,LI-610} = 1.103\chi_{v,EC150} - 1.06$	$\chi_{v,LI-610} = 1.049\chi_{v,EC150} - 0.20$

3.2. Standard calibration versus LGR

We evaluated the 2nd and 3rd generation hygrometers versus the first principles based LGR. The results in Table 2 show that for all fast-response hygrometers the R^2 of the linear regression was at least 0.98 and the RMSE error was less than 0.20 mmol mol^{-1}, with the best being the relationship between the LI-7200 and the LGR (Fig. 3). When the LGR was predicted by higher order polynomials (i.e., first summation in Eq. (1)), there was minimal increase in the R^2 or decrease in the RMSE, with the largest change occurring in the EC150, and the LI-7200 behaving like the others (i.e., the LI-7200 response was reasonably linear). When a time component was included (i.e., second summation in Eq. (1)), the smallest changes in R^2 and RMSE were with the LI-7200 (i.e., the LI-7200 response was reasonably stable over time).

Based on the calibration drifts reported in Section 3.1, we considered that some of the standard calibrations might be inaccurate. Thus, we adjusted the standard calibrated data relative to the LGR. Here we focused on the standard calibration without the zero-air as discussed above. Considering all hygrometers had an $R^2 > 0.98$ when using a simple linear response with the LGR, we only focus on the first order coefficient c_1 (i.e., the slope in Eq. (1) and Table 2) as it directly multiplies into the covariance between vertical wind velocity and water vapor (either molar mixing ratio or mass density) which manifests into water fluxes and impacts the energy balance. The smallest slope was for the LI-7500 (i.e., 0.975), indicating that the standard calibrated LI-7500 might measure fluxes that are 2.5% too high. The largest slope was for the LI-7000 (i.e., 1.166) indicating that the standard calibrated LI-7000 might measure fluxes that are 16.6% too small.

3.3. Piecewise calibration

The piecewise calibration was completed for all fast-response hygrometers, excepting the LI-7200. Though a goal of this study was not to condone a specific hygrometer as a standard, piecewise calibration requires a reference hygrometer. Because the LI-7200 had the best response relative to the LGR and the largest valid sample size, we selected the LI-7200 as the reference. The LI-7000 was generally the

Table 2
Results of regression analysis comparing vapor concentration (χ_v) from different hygrometers using the standard calibration (without zero-air) to the first-principles based LGR where $\chi_{v,LGR} = c_1 * \chi_{v,hygrometer} + c_0$. For regressions predicting $\chi_{v,LGR}$ from higher order polynomials of $\chi_{v,hygrometer}$ (n^{th} order) and time (m^{th} order), i.e. Eq. (1), the degree of each polynomial was selected by maximizing the coefficient of determination (R^2) and minimizing the root-mean-square-error (RMSE). The Δ values are the changes in the higher order regression statistics relative to the 1st order regression. Results for the Lyman-α and KH2O are not included because of severe instrument drift. Data collected at GLEES from 11 September to 2 October 2015.

				1st order regression vs. LGR				n^{th} order regression vs. LGR			1st order regression vs. LGR, m^{th} order regression vs. time		
Hygrometer	Generation	Type	# of samples	R^2	RMSE	c_1 (unitless)	c_0 (mmol mol^{-1})	n	ΔR^2	ΔRMSE	m	ΔR^2	ΔRMSE
LI-7000	2nd	fast response, closed-path	455	0.9908	0.1552	1.166	−0.04	9	0.0007	−0.0053	7	0.0042	−0.0397
LI-7500	2nd	fast response, open-path	480	0.9928	0.1374	0.975	0.03	6	0.0004	−0.0026	7	0.0027	−0.0268
LI-7200	3rd	fast response, closed-path	455	0.9953	0.1105	1.111	0.22	9	0.0006	−0.0058	8	0.0005	−0.0046
EC155	3rd	fast response, closed-path	96	0.9928	0.1262	1.066	0.58	7	0.0010	−0.0053	6	0.0034	−0.0281
EC150	3rd	fast response, open-path	96	0.9827	0.1955	1.091	0.20	8	0.0044	−0.0200	6	0.0127	−0.0882
Met-One 083D	n/a	slow response, RH chip	486	0.9946	0.1223	1.037	0.32	8	0.0002	−0.0013	7	0.0016	−0.0180

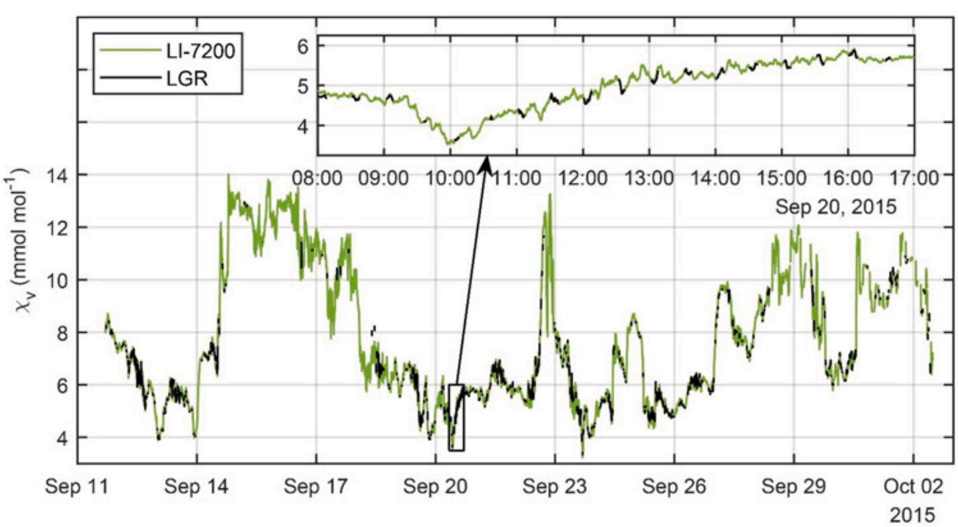

Fig. 3. Water vapor concentration (χ_v) measured with the LI-7200 and the first principles based LGR, which is used to assess hygrometer calibration accuracy and stability over time (Table 2). The LI-7200 data were processed with an exponential moving average to mimic the slow response of the LGR; the 1st order regression coefficients from Table 2 are applied here to overlay the plots, i.e., Eq. (1) with $N = 1$ and $M = 0$. The LGR data are sporadic and occur for 5–7 min of every half hour; the instrument was configured to sample along a vertical profile on the GLEES scaffold, with the data here sampled at the top.

most similar to the LI-7200, with standard calibrated concentrations of both sensors generally falling on the 1:1 line (Fig. 4a). Using a threshold of 0.00008 mmol mol^{-1} for the piecewise calibration algorithm, 91.6% of the data was identified into 120 blocks (Fig. 4c-e) resulting in a RMSE of 0.0184 mmol mol^{-1} between the LI-7000 and LI-7200 (Fig. 4b). The KH2O experienced severe instrument drift, causing a poor relationship between the raw KH2O and the LI-7200 (Fig. 5a). Using a threshold of 0.002 g m^{-3}, 79.2% of the data was assigned into 334 blocks (Fig. 5c-e) resulting in a RMSE of 0.0143 g m^{-3} between the KH2O and LI-7200 (Fig. 5b). For the other hygrometers, the thresholds were 0.00012 g m^{-3} for the Lyman-α (254 blocks, 76.6% of the data, 0.0111 g m^{-3} RMSE), 0.00006 g m^{-3} for the LI-7500 (113 blocks, 65.1% of the data, 0.0104 g m^{-3} RMSE), 0.00012 mmol mol^{-1} for the EC155 (38 blocks, 82.5% of the data, 0.0171 mmol mol^{-1} RMSE), 0.000035 g m^{-3} for the EC150 (69 blocks, 73.9% of the data, 0.0059 g m^{-3} RMSE), 0.00006 g m^{-3} for the secondary calibration of the LI-7500 relative to the EC155 (64 blocks, 54.5% of the data, 0.0077 g m^{-3} RMSE), and 0.00003 g m^{-3} for the secondary calibration of the EC150 relative to the EC155 (67 blocks, 50.2% of the data, 0.0057 g m^{-3} RMSE). In each case, the listed thresholds were chosen to optimize the overall performance of the piecewise calibration, based on the number of blocks identified, the percent of data that was incorporated into a block, and the RMSE of the fit with the LI-7200. Tradeoffs of these choices are discussed in section 4.4. To better evaluate the suitability of piecewise calibration algorithm for slow-response data, we repeated the procedure on a dataset of 5-minute medians, where in general, a lower threshold was required to achieve similar performance. These thresholds were 0.00008 g m^{-3} for the Lyman-α (106 blocks, 67.9% of the data, 0.0085 g m^{-3} RMSE), 0.00012 g m^{-3} for the KH2O presented in the supplementary material Figure S1 (103 blocks, 69.7% of the data, 0.0104 g m^{-3} RMSE), 0.00006 mmol mol^{-1} for the LI-7000 presented in the supplementary material Figure S2 (79 blocks, 86.4% of the data, 0.0140 mmol mol^{-1} RMSE), 0.00003 g m^{-3} for the LI-7500 (63 blocks, 62.4% of the data, 0.0067 g m^{-3} RMSE), 0.00009 mmol mol^{-1} for the EC155 (33 blocks, 87.5% of the data, 0.0133 mmol mol^{-1} RMSE), 0.000025 g m^{-3} for the EC150 (26 blocks, 59.6% of the data, 0.0050 g m^{-3} RMSE), 0.000035 g m^{-3} for the secondary calibration of the LI-7500 relative to the EC155 (36 blocks, 57.5% of the data, 0.0048 g m^{-3} RMSE), and 0.0000175 g m^{-3} for the secondary calibration of the EC150 relative to the EC155 (32 blocks, 52.0% of the data, 0.0039 g m^{-3} RMSE).

Fig. 6 shows the complete piecewise calibrated dataset based on the 5-minute quantiles used in this study. Not including the EC155 and EC150, the other hygrometers were successfully calibrated for most of the study period, except for the Lyman-α, which inadvertently filled with

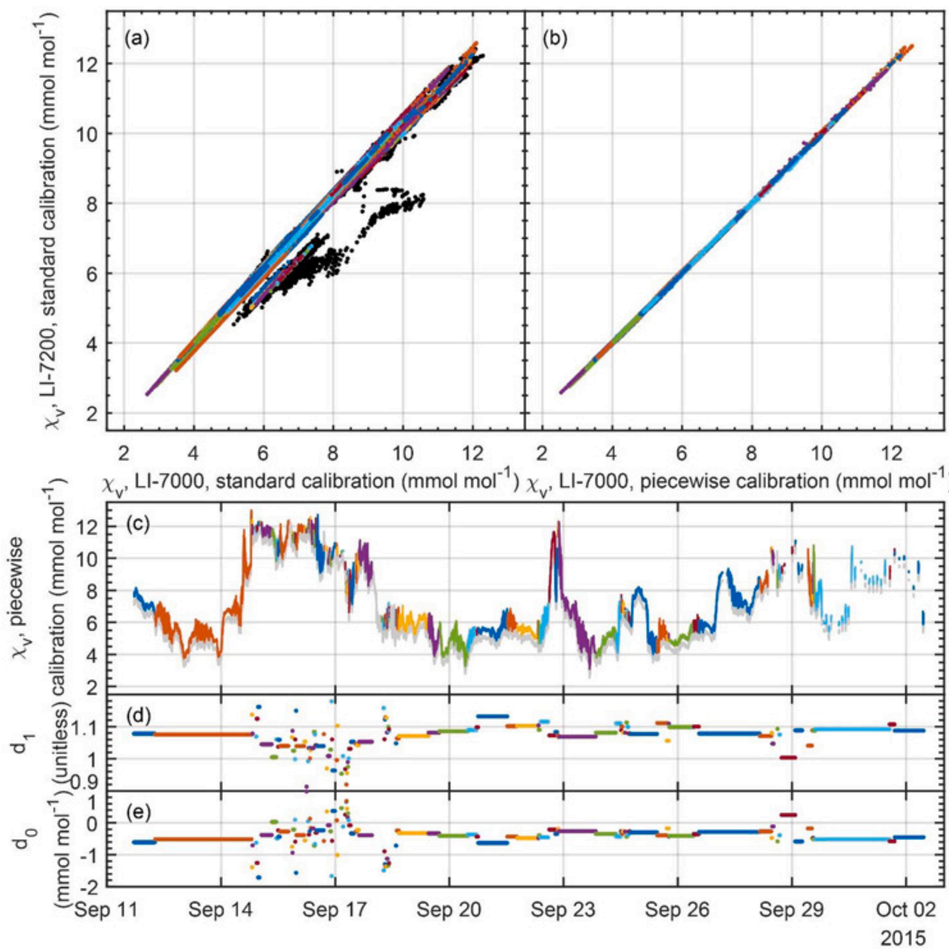

Fig. 4. The piecewise calibration of the LI-7000 hygrometer water vapor concentration (χ_v) collected at GLEES from 11 September to 2 October 2015. (a) First, blocks of time (i.e., various colors) are identified where the LI-7000 (i.e., standard calibration) and the LI-7200 (i.e., standard calibration) have a consistent linear response among their 5-minute quantiles (5, 10, 25, 50, 75, 90, and 95%); data not assigned to any block are in black. (d, e) Next, linear regression is performed for each block using all quantiles to estimate the regression coefficients for the slope (d_1) and intercept (d_0); variation and drifts in these parameters are generally attributable to uncertainty in either hygrometer. (b, c) Finally, the coefficients are applied to the LI-7000. In (c), the y-axis is referenced to the LI-7200 (gray) with the LI-7000 slightly offset vertically for better visibility.

water during a rainstorm storm and was repaired about a week later. The inset of Fig. 6 shows the piecewise calibrated water vapor concentrations for all seven hygrometers; most of the missing data at the end are due to electrical problems resulting from another rainstorm. The inset demonstrates how the piecewise calibration not only corrects for slow response instrument drift, but it retains the turbulent dynamics in variance among the sensors by fitting the quantiles.

3.4. Differences in water vapor fluxes

We obtained water vapor fluxes throughout most of the study period for the KH2O, LI-7000, LI-7500, and LI-7200 when fluxes were calculated with the ATI 3Vx-probe; the Lyman-α was offline for about a week due to sensor issues (Fig. 7). The inset of Fig. 7 shows the water vapor fluxes for all seven hygrometers when the EC155 and EC150 were installed and when the co-located CSAT3 anemometer could be used for covariance calculations. The appendix with Figs. A1–7 displays the signal from each hygrometer during a twenty-minute period with relatively large water vapor fluxes and while all sensors had successful piecewise calibrations; each is shown as a time series, in the frequency domain, and as a wavelet spectrogram. The appendix also includes the covariance/cospectra between the vertical wind velocity and water vapor concentration. Subjectively, the LI-7200 and EC150 have clean and relatively ideal frequency responses (Figs. A5b-c and A7b-c). The Lyman-α has a noise peak at 5.5–7 Hz, an elevated noise floor above 4 Hz, and nearly a dozen smaller noise peaks distributed between 0.5–5 Hz. (Fig. A1b-c), the KH2O has elevated noise above 1 Hz (Fig. A2b-c), the LI-7000 has a dampened response above 1.5 Hz (Fig. A3b-c, note, this is an older unit), the LI-7500 has a noise peak at 5 Hz and an elevated noise floor above 5 Hz (Fig. A4b-c), and the EC155 has noise peaks at ~ 0.7, ~ 1.8–1.9, and ~ 4.3 Hz and a dampened response above 0.2 Hz (Fig. A4b-c).

The results of the Bayesian analysis comparing the water vapor fluxes are shown in Fig. 8. Regardless of whether the zero-air point is included (Fig. 8a-b versus c-d) or if the fluxes are calculated with the CSAT3 or the ATI 3Vx-probe (Fig. 8a, c versus b, d), the differences between water vapor fluxes were larger with the standard calibration than for any other calibration method. The maximum difference of 15.9% occurred between the LI-7500 and LI-7000 when using the standard calibration without zero-air (Fig. 8a). The standard deviations of% differences among the 2nd and 3rd generation sensors using the standard calibration without zero-air were 6.0 and 5.7% for fluxes calculated with the CSAT3 and ATI 3Vx-probe, respectively; when the zero-air was included these standard deviations were 4.9 and 4.1%. When the standard calibration is adjusted using the regression results relative to the LGR, the differences are reduced (Fig. 8e-f versus a-d); the maximum difference of 8.9% occurred between the EC150 and LI-7500 when using the ATI 3Vx-probe (Fig. 8f), which is less than the smallest range using the standard calibration (Fig. 8a-d). In this case the standard deviations of the% differences were 2.4 and 3.3% for the two anemometers, respectively. The piecewise calibration resulted in smaller differences, with the maximum difference between 2nd and 3rd generation sensors of 8.2% occurring between the EC150 and LI-7500 when using the ATI 3Vx-probe (Fig. 8h). The smallest range between 2nd and 3rd generation sensors occurred with the piecewise calibration when using the CSAT3 (Fig. 8g) where a 5.4% difference was detected between the EC150 and LI-7500. For the piecewise calibration, the standard deviation of the% differences among the 2nd and 3rd generation sensors were 2.3 and

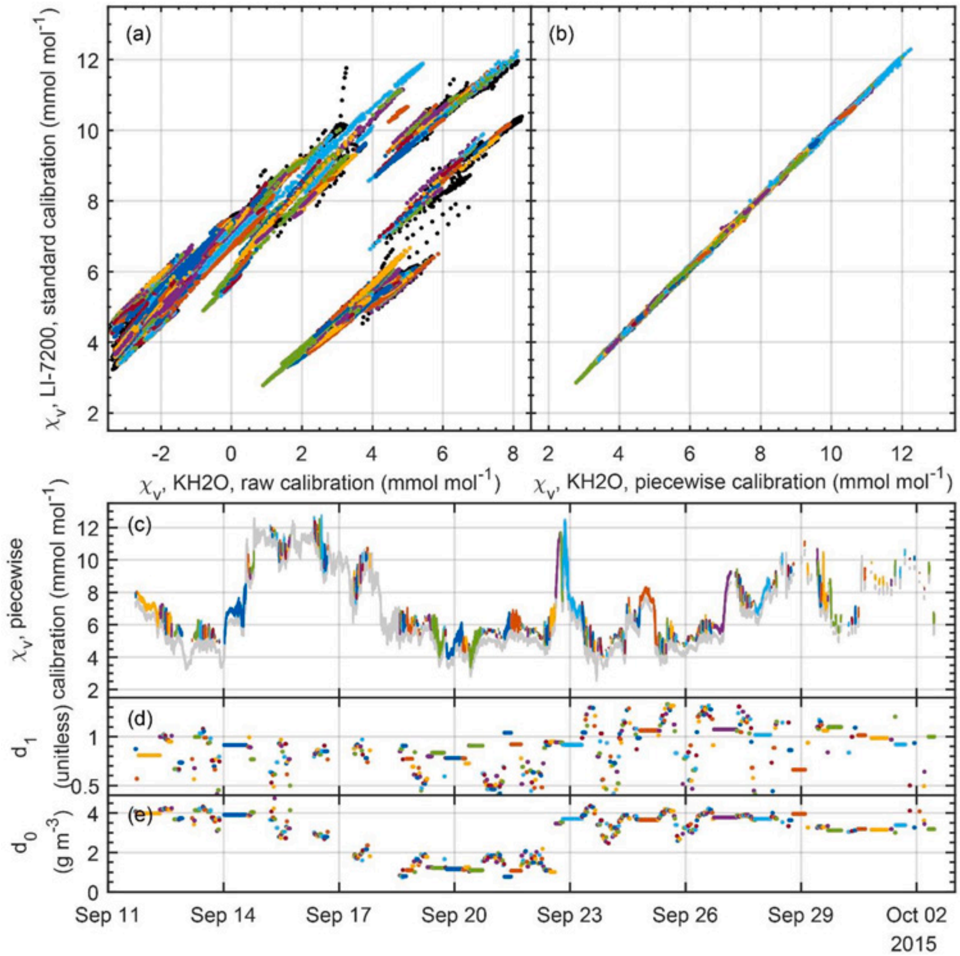

Fig. 5. The piecewise calibration of the KH2O hygrometer water vapor concentration (χ_v) collected at GLEES from 11 September to 2 October 2015. (a) First, blocks of time (i.e., various colors) are identified where the KH2O (i.e., raw calibration) and the LI-7200 (i.e., standard calibration) have a consistent linear response among their 5-minute quantiles (5, 10, 25, 50, 75, 90, and 95%); data not assigned to any block are in black. (d, e) Next, linear regression is performed for each block using all quantiles to estimate the regression coefficients for the slope (d_1) and intercept (d_0); variation and drifts in these parameters are generally attributable KH2O because far less drift is possible in the LI-7200 (Fig. 4d-e). (b, c) Finally, the coefficients are applied to the KH2O. In (c), the y-axis is referenced to the LI-7200 (gray) with the KH2O slightly offset vertically for better visibility.

3.0% for the two anemometers, respectively. The median σ, i.e., the standard deviation of the prediction, had an average across all analyses (Fig. 8a-l) of 10.2 mmol $m^{-2} s^{-1}$ and ranged from 7.1 to 14.2 mmol $m^{-2} s^{-1}$. The results for comparing water vapor fluxes using a piecewise calibration based on 5-minute medians were similar (Fig. S3). The standard deviations of% differences among the 2nd and 3rd generation sensors using the standard calibration ranged from 5.3 to 6.4% which was reduced with the piecewise calibration to 2.8% using the CSAT3 and 3.4% using the ATI 3Vx-probe. The median σ averaged across all analyses (Fig. S3a-l) was 12.2 mmol $m^{-2} s^{-1}$ and ranged from 10.2 to 15.5 mmol $m^{-2} s^{-1}$.

The 1st generation hygrometers with the piecewise calibration tended to measure fluxes within the range of the 2nd and 3rd generation sensors (Fig. 8g-h) with the notable inconsistency of the KH2O, where fluxes measured using the CSAT3 were larger than from any other sensor (i.e., 4.0% larger than the EC150 in Fig. 8g); in analyses using the piecewise calibration based on 5-minute medians, all 1st generation sensors measured fluxes within the range of the others (Fig. S3g-h). Omitting the O_2 correction for the 1st generation sensors decreased the water vapor fluxes, notably shifting the KH2O from the highest fluxes to among the lowest (Fig. 8i-j versus g-h). Omitting the tube attenuation correction for the closed-path sensors had a less dramatic effect, though it did reduce the EC155 fluxes by about 1.4% relative to the others (Fig. 8k versus i).

We investigated potential errors in calculating fluxes between sensors originally sampled at 10 Hz and 20 Hz, where our convention was to resample the sonic anemometer to match the sample rate of the hygrometer. We estimated this error by examining covariances between the co-located ATI 3Vx-probe and LI-7500 which were both sampled at 20 Hz. We calculated covariances with both at 20 Hz, with the ATI 3Vx-probe down-sampled to 10 Hz then up-sampled via linear interpolation back to 20 Hz, and finally both down-sampled to 10 Hz. The effect of up-sampling the sonic anemometer to 20 Hz caused a dampening in high frequencies of the cospectra above 5 Hz (Fig. A4f). The differences in covariances caused by up-sampling to 20 Hz versus down-sampling to 10 Hz was 0.1–0.2% with a maximum $< 1\%$. Thus, we do not consider the mismatch in sampling rate to substantially effect our results.

We compared the averaged 5-minute fluxes to the 30-minute US-GLE data. For 825 pairs of H data the RMA slope was 0.98 with an intercept of −0.3 W m^{-2}, R^2 of 0.9948, and RMSE of 11.4 W m^{-2}. For 820 pairs of F_v data the RMA slope was 0.92 with an intercept of 0.036 mmol $m^{-2} s^{-1}$, R^2 of 0.9372, and RMSE of 0.306 mmol $m^{-2} s^{-1}$. For the water vapor flux comparison there were six notable outliers; with their omission the RMA slope was 0.96 with an intercept of 0.016 mmol $m^{-2} s^{-1}$, R^2 of 0.9617, and RMSE of 0.229 mmol $m^{-2} s^{-1}$.

3.5. Differences due to vertical separation of anemometers and scalar sensors

The effect of vertical separation between the anemometers and scalar sensors is demonstrated with sensible heat flux measurements in Table 3. When using the vertical wind velocity from 25.2 m height and switching the sonic virtual temperature from 25.2 m down to 22.65 m, i.e., scalar sensor below the anemometer, the sensible heat flux increased 4.0%. When using the vertical wind velocity at 22.65 m height and switching the sonic virtual temperature from 22.65 up to 25.2 m, i.e., scalar sensor above the anemometer, the sensible heat flux changed by −1.6%. These results do account for the horizontal separation between

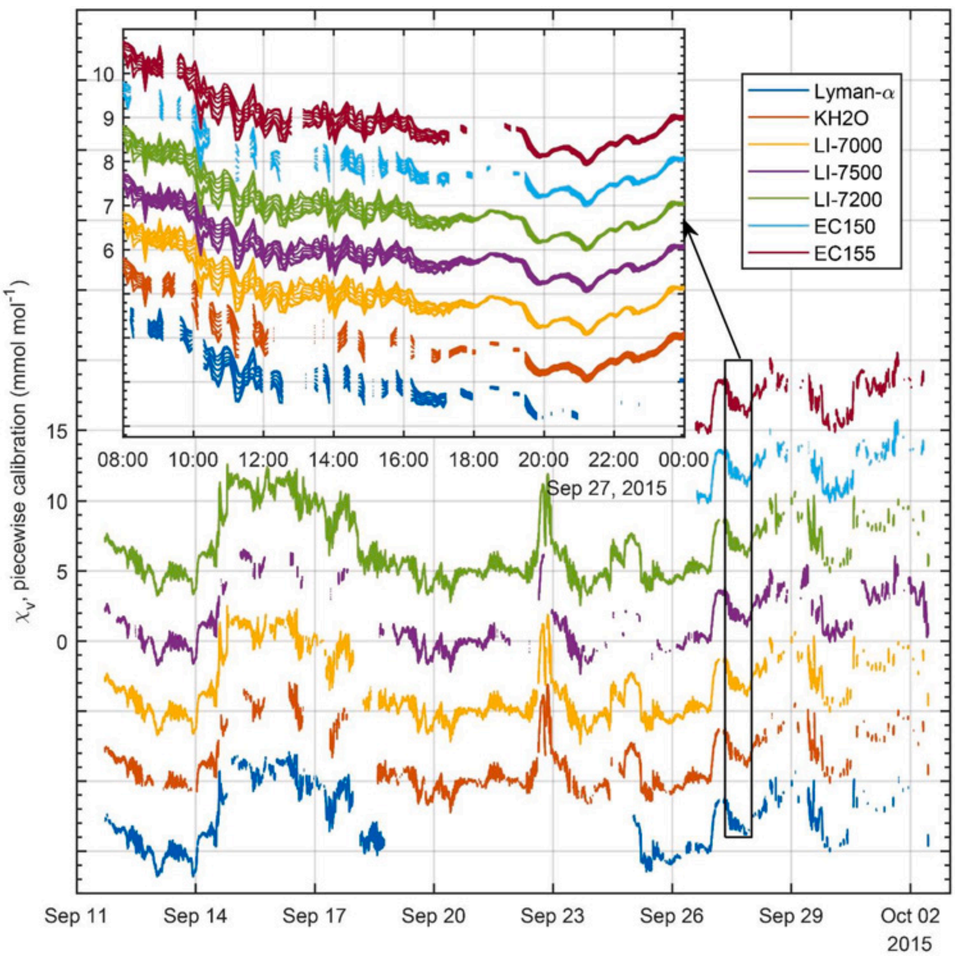

Fig. 6. Stacked water vapor concentration (χ_v) collected at GLEES from 11 September to 2 October 2015 from hygrometers using the piecewise calibration, which uses the LI-7200 (i.e., standard calibration) as the reference and matches the other hygrometers using their 5-minute quantiles (5, 10, 25, 50, 75, 90, and 95%, shown as multiple lines). As shown in the inset, the piecewise calibration simultaneously accounts for the slow drift in each sensor along with their fast response variation as the spread between lines for each quantile changes from wider to narrower. The y-axis is referenced to the LI-7200 with the other hygrometers stacked vertically for better visibility.

the sonic anemometer and hygrometer (i.e., ~ 1.4 m as shown in Fig. 1), but do not account for vertical separation (i.e., ~ 2.5 m as shown in Fig. 1) or for any effects on temperature measurement due to changing between the CSAT3 and ATI 3Vx-probe. Evaluating the methods of Kristensen et al. (1997) at the same heights results in changes of −1.1% and −10.1%, respectively. Our results average 7% higher than Kristensen et al. (1997), though in both sets of values the effect of putting the scalar above the anemometer resulted in significantly lower fluxes than having the scalar measurement below. Therefore, caution should be used when comparing fluxes from sensors that were not co-located. Comparisons using the ATI 3Vx-probe should be valid in a relative sense because biases due to separation should be consistent between the hygrometers. Comparisons using the LI-7500 are more problematic because of its dissimilar location from the sonic anemometer, regardless of the CSAT3 or ATI 3Vx-probe. Thus, in the following sections we interpret our results with and without the inclusion of the LI-7500.

3.6. Differences in energy balance closure

Unlike the previous results, analysis of the energy balance closure was highly sensitive to unequal sample sizes amongst the various hygrometers. Haslwanter et al. (2009) had a similar occurrence, where their difference in energy balance closure between two hygrometers was 2% when using paired samples and 5% when using unequal sample sizes. We propose the reason for this is that unlike the previous analyses where the predictions are a function of the "baseline" which is conceptually the average flux from all the hygrometers, predictions for the energy balance are based on an independent measurement of the available energy. Considering that energy balance closure his highly variable and typically has a diurnal cycle (Reed et al., 2018), with unequal sample sizes it is probable that the available energy will predict one set of data differently than another. Unfortunately, the largest sample size with complete energy balance data for all seven hygrometers was 175 5-minute periods (Fig. 9a-c). While there were no statistically significant differences, the largest discrepancy in the closure ratio was 0.022 between the EC150 and LI-7000 using the standard calibration (Fig. 9a), whereas the largest discrepancy using the piecewise calibration was 0.019 between the Lyman-α and KH2O (Fig. 9b). The largest sample size with complete energy balance data for two hygrometers using the CSAT3 was 755 5-minute periods using the LI-7000 and LI-7200 (Fig. 9d, f, h) where the closure ratio had a discrepancy of 0.008 with the standard calibration (Fig. 9d) which decreased to 0.003 with the piecewise calibration. Using the same sample size but calculating fluxes with the ATI 3Vx-probe had minimal impact on the discrepancies between the LI-7000 and LI-7200 (Fig. 9d, f, h versus e, g, i) but did increase the energy balance closure ratio by an average of 0.048. The largest sample size with complete energy balance data for two hygrometers using the ATI 3Vx-probe was 4488 5-minute periods using the LI-7000 and LI-7200 (Fig. 9j-l) where the largest discrepancy in the closure ratio was 0.010 for the standard calibration (Fig. 9j) which decreased to 0.004 for the piecewise calibration (Fig. 9k). For this subset of data, the average energy balance closure ratio was 0.859 using the piecewise calibration. Considering the piecewise calibration is referenced to the standard calibrated LI-7200 which underestimated the LGR by ~ 11% (Table 2), adjusting the piecewise calibration to reference the LGR increased the average closure ratio to 0.882. The results for estimating energy balance closure using a piecewise calibration based on 5-minute medians were similar (Fig. S4), with the main difference being

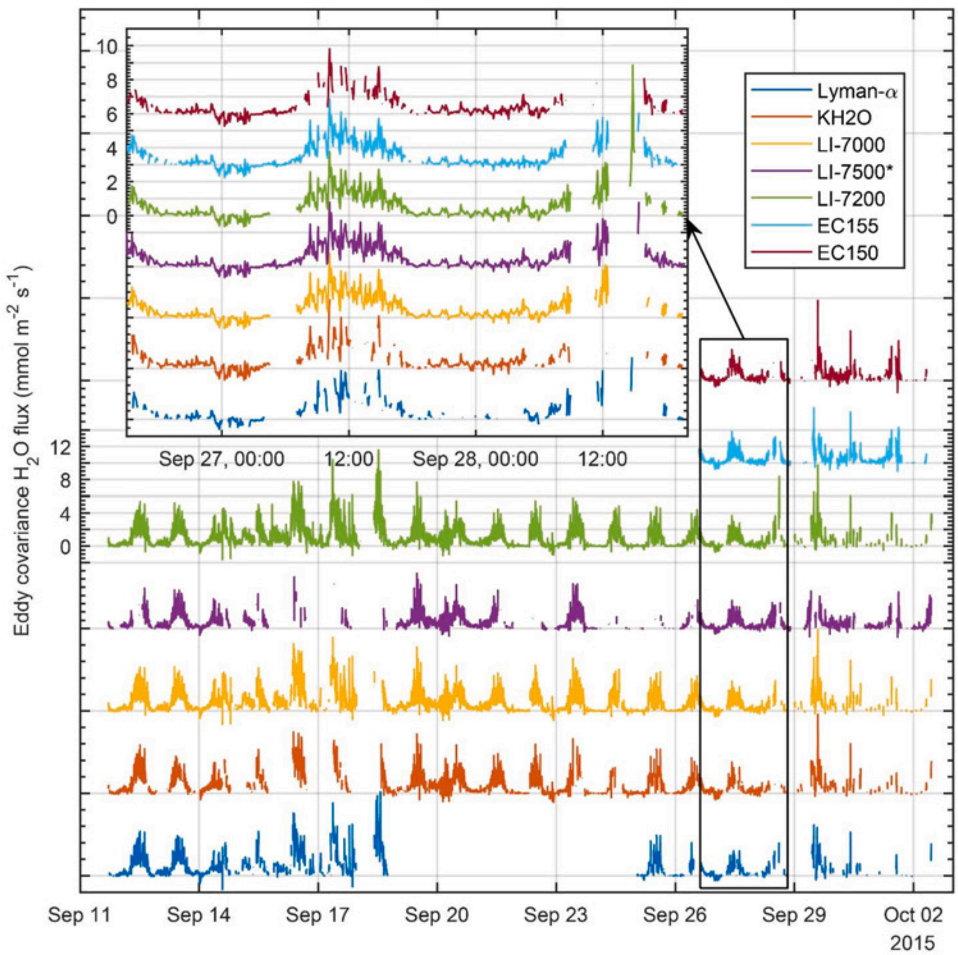

Fig. 7. Stacked 5-minute water vapor flux collected at GLEES from 11 September to 2 October 2015 from hygrometers using the piecewise calibration and calculated with the ATI 3Vx-probe located ~ 2.5 m below (Fig. 1, *also the location of the LI-7500). In the inset all seven hygrometers are present and fluxes were calculated with the co-located CSAT3 anemometer. The y-axis is referenced to the LI-7200 with the other hygrometers stacked vertically for better visibility.

the analyses using all seven hygrometers (Fig. S4a-c). The largest sample size with complete energy balance data for all hygrometers was 236 5-minute periods and resulted to significantly lower closure ratios, ~ 0.79, than the corresponding analyses with the piecewise calibration based on quantiles, ~ 0.91 (Fig. S4a-c vs. Fig. 9a-c).

4. Discussion

4.1. The effect of calibration on measured water vapor fluxes

In our study, the uncertainty in calibration caused the greatest variation among measured eddy covariance H_2O fluxes. We attribute this to two concerns, the ability to perform an accurate standard calibration at the field site over the full range of humidity and the stability of instruments to maintain their calibration.

It can be difficult to conduct an accurate calibration of most hygrometers over the full range of naturally occurring humidity at a field site. At our Rocky Mountain site, the dew point is often < 0 °C, both during arid summers and cold winters making it impossible to use a dew point generator to produce ambient water vapor concentrations. One option is to generate a series of dew points above 0 °C, then to interpolate the calibration into the lower range of humidity that naturally occurs. For the 2nd and 3rd generation hygrometers (i.e., all but the Lyman-α and KH2O), these multipoint calibrations without the zero-air data fit on a near-perfect straight line (i.e., blue dots in Fig. 2a for the EC150, RMSE of linear regression ≤ 0.034 mmol mol^{-1} for all hygrometers). The benefit of the zero-air is it provides a reference point lower than ambient conditions, thus better constraining the extrapolation of the calibration into ambient conditions that are otherwise untestable. One negative is it relies on a completely different methodology (i.e., a zero-air tank, etc., versus a dew point generator) that requires complete purging of all vapor within the tubing and sample cell. Another is the zero point is at the extreme end of what is typically a non-linear absorption response to water vapor. For example, the LI-7000 uses a factory calibrated 3rd order polynomial to linearize the response; if a discrepancy develops between the sensor response and the factory calibration then forcing a calibration through zero can have non-linear effects in the range of ambient measurements. As we found, calibration regressions can fit substantially worse when the zero-air tank is included. This was most prominent for the EC155 and EC150 where the RMSE increased by a factor of ~ 30 × to 60 × and the regression slopes decreased by −5% (Fig. 2a for the EC150). The zero-air had a less extreme effect on the LI-COR hygrometers, which experienced ±2% changes in their calibration slopes.

Another problem is the stability of the calibrations. Only the LI-COR hygrometers received multiple standard calibrations, with the LI-7000 and LI-7200 experiencing −6 to −9% decreases in their calibration slopes after one week (Table 1). A potential explanation for this change could be the atmospheric conditions on 17 September 2015 which limited reliable dew point generation to between 1 and 4 °C. Another explanation is that both the LI-7000 and LI-7200 are closed-path analyzers, such that it is difficult to generate enough vapor at the flow rate required by the hygrometer. We observed that while the dew point generator was connected in series with the closed-path analyzers the

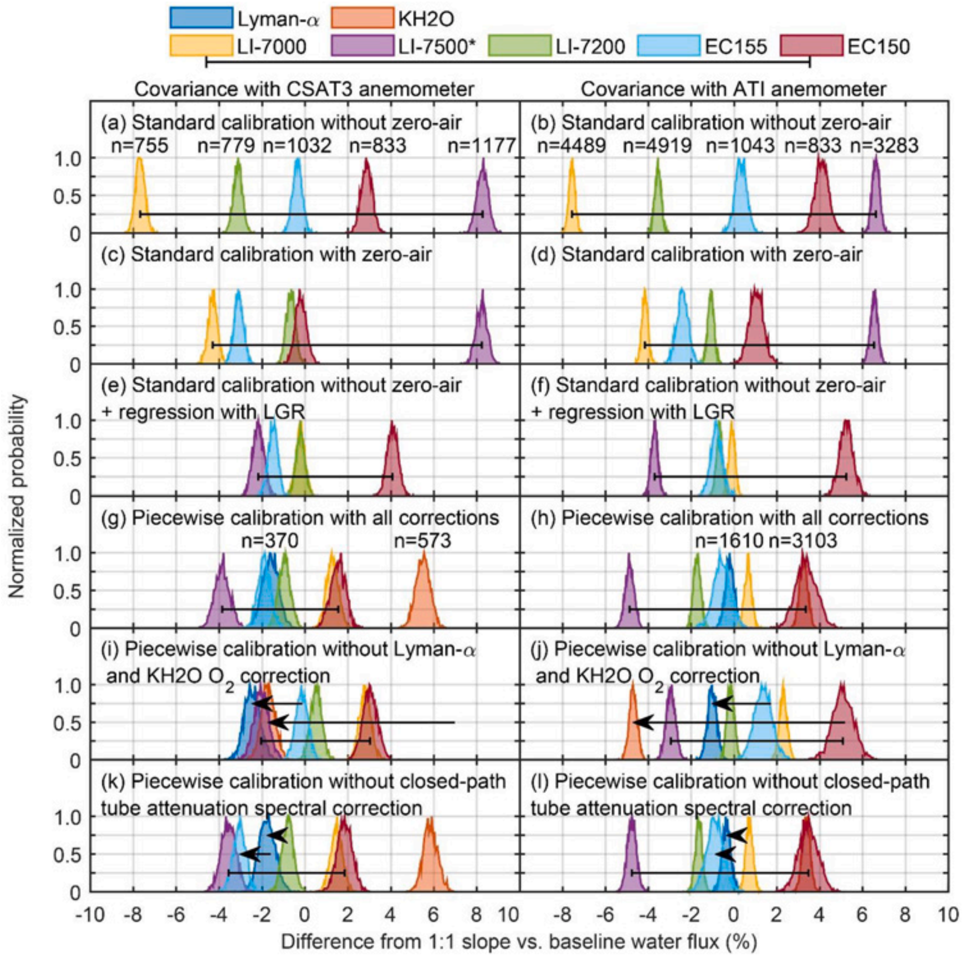

Fig. 8. A comparison of eddy covariance H_2O fluxes from hygrometers processed with (a, b) the standard calibration without zero-air, (c, d) with zero-air, (e, f) adjusted relative to the LGR (i.e., Table 2), and (g, h) the piecewise calibration (i, j) without the Lyman-α and KH2O oxygen correction or (k, l) without the closed-path tube attenuation spectral correction; all datasets correspond to the piecewise calibration using 5-minute quantiles. Fluxes were calculated with (a, c, e, g, i, k) the co-located CSAT3 anemometer or (b, d, f, h, j, l) the ATI 3Vx-probe ~ 2.5 m below (Fig. 1). The posterior distributions are Bayesian process parameters from the statistical analysis predicting water vapor fluxes from the percent difference from the "baseline" water flux (i.e., predicted 5-minute "baseline" fluxes represent average conditions and are also process parameters with posterior distributions). The Lyman-α and KH2O are not included in the standard calibration because of severe instrument drift. Brackets denote the range in response between 2nd and 3rd generation hygrometers. Arrows show the relative shift in position when corrections are omitted. *The LI-7500 was co-located with the ATI 3Vx-probe.

Table 3

The effect of vertical separation of anemometers/scalar sensors demonstrated using sensible heat flux calculated with combinations of vertical wind (w) and virtual temperature (T_s) from the CSAT3 and ATI 3Vx-probe (Fig. 1). The differences are Bayesian process parameters with posterior distributions (95% Bayesian credible intervals in parentheses) estimated from the statistical analysis predicting sensible heat fluxes as the percent difference from the "baseline" heat flux (i.e., predicted 5-minute "baseline" fluxes represent average conditions and are also process parameters with posterior distributions). Number of samples = 1614. Data collected at GLEES from 11 September to 2 October 2015.

Vertical Separation	Height (m)		Sensor		Difference from 1:1 slope vs. baseline sensible heat flux (%)	Change in sensible heat flux due to vertical separation (%)	Predicted flux loss due to vertical separation from Kristensen et al. (1997) (%)	Median spectral correction for horizontal separation (%)
	w	T_s	w	T_s				
Co-located	25.2	25.2	CSAT3	CSAT3	−3.6 (−3.9 to −3.2)	n/a	n/a	n/a
Scalar below sonic anemometer	25.2	22.65	CSAT3	ATI 3Vx-probe	0.5 (0.2 to 0.8)	4.0 (3.5 to 4.6)	−1.1	9.8
Co-located	22.65	22.65	ATI 3Vx-probe	ATI 3Vx-probe	2.3 (2.0 to 2.7)	n/a	n/a	n/a
Scalar above sonic anemometer	22.65	25.2	ATI 3Vx-probe	CSAT3	0.7 (0.4 to 1.1)	−1.6 (−2.2 to −1.1)	−10.1	11.2

flow rate was reduced to around the 2 LPM limit of the LI-610 and that the internal pressure of the closed-path analyzers fell around −2 kPa. A problem is the calibration is occurring under lower flows and lower pressures than in normal operations; plus, at these flow rates the dew point generator may not be able to produce vapor at the concentrations prescribed by the instrument set point. On 17 September, we conducted an abbreviated calibration while the pumps on the LI-7000 and LI-7200 were bypassed, which tended to overpressure the analyzer cells by ~ 1 kPa and increased the calibration slopes by ~ 10% and ~ 4%, respectively, further illustrating the difficulty in calibrating closed-path sensors under normal operating conditions. Finally, the standard calibration of a vapor concentration requires knowledge of the internal pressure of the dew point generator. In our calculations we estimated this as the ambient pressure from the GLEES AmeriFlux site, which could easily be ±2 kPa different from the actual pressure within the LI-610 based on the considerations above; we estimate this could change the calibration slopes by ±3%. While it is impossible to know from the repeated calibrations whether the LI-7000 and LI-7200 truly drifted over

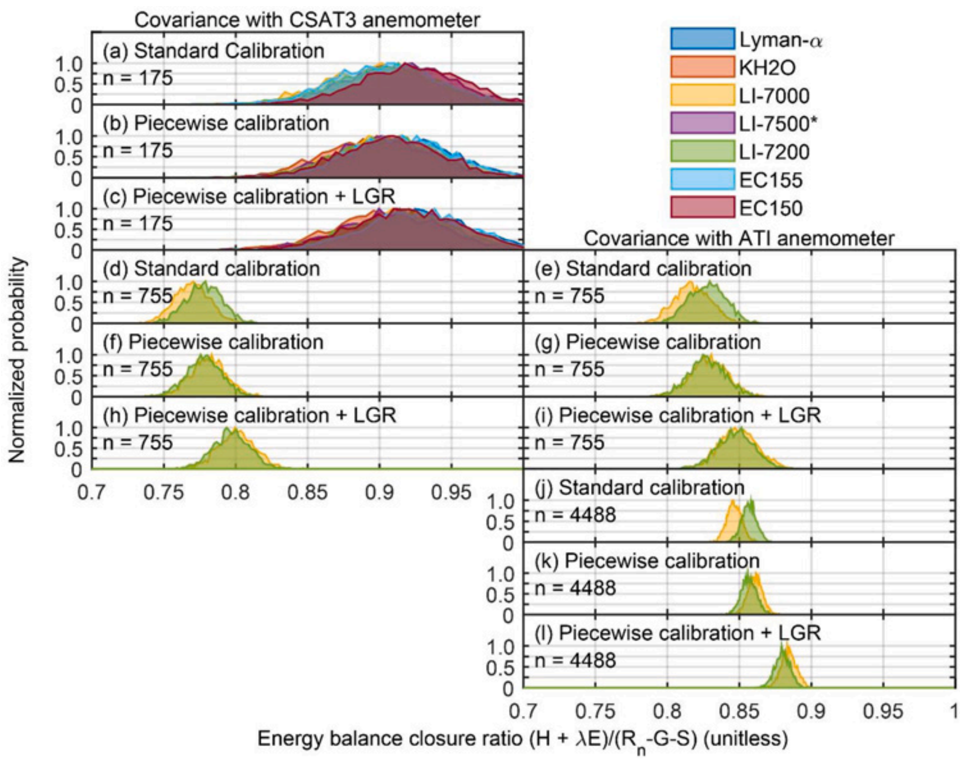

Fig. 9. Energy balance closure ratio evaluated when latent heat (λE) is calculated from hygrometers processed with (a, d, e, j) the standard calibration without zero-air, (b, f, g, k) the piecewise calibration, and (c, h, i, j) the piecewise calibration adjusted relative to the LGR; all datasets correspond to the piecewise calibration using 5-minute quantiles. (a, b, c) $n = 175$ is the maximum sample size of 5-minute data containing values from all seven hygrometers, (d, f, h) $n = 755$ is the largest dataset with both LI-7000 and LI-7200 fluxes calculated with the co-located CSAT3 anemometer and is directly comparable with (e, g, i) fluxes calculated with the ATI anemometer ~ 2.5 m below (Fig. 1), and (j, k, l) $n = 4488$ is the largest dataset with both LI-7000 and LI-7200 fluxes. The posterior distributions are Bayesian process parameters from the statistical analysis predicting turbulent fluxes ($H + \lambda E$) from available energy ($R_n - G - S$). (c, h, i, j) Because the piecewise calibration is referenced to the LI-7200, to adjust all λE to be relative to the LGR, only the LI-7200 vs. LGR regression coefficient (i.e., × 1.111, Table 2) needs to be applied.

time or if there were complications due to methodology, the overall effect is the same: fluxes calculated based on one calibration versus the other would be 6–9% different. A final explanation for the instability of calibrations concerns open-path analyzers where it is often difficult to control the climate in the measurement volume, especially with the LI-7500 where the calibration chamber loosely seals around the upper and lower windows. Though, in our study the LI-7500 performed well in its repeated calibrations (Table 1). Ultimately, we found that for the standard calibration the average absolute change in calibration slope due to the stability (i.e., repeatability) was ~ 6%, which was 3 × larger than the average absolute change in calibration slope due to the inclusion/exclusion of the zero-air tank which was ~ 2% (Table 1).

A different way to quantify the stability of hygrometer calibrations over time is to compare them to a stable reference or to identify the drift in regression coefficients created by the piecewise calibration. We considered the LGR to be a stable reference because its calibration is based on first principles, and it was housed inside the climate-controlled shelter at the base of the scaffold. The LI-7200 had the smallest change in the regression analysis versus the LGR over time (Table 2, Fig. 3). The piecewise calibration is useful to detect drifts between two sensors over time, which in the case of the LI-7000 versus LI-7200 (Fig. 4d-e), the drift might be more attributable to the LI-7000 because the LI-7200 was shown to be more stable relative to the LGR.

Ultimately, we consider the piecewise calibration to have the minimum calibration uncertainty between hygrometers. Therefore, differences in fluxes using the standard calibration, etc., can be compared to the piecewise calibration to determine the uncertainty due to calibration methodology. We found the uncertainty in water vapor fluxes among the 2nd and 3rd generation sensors was greatest if hygrometers only received the standard calibration, regardless of the inclusion of zero-air (Fig. 8a-d, maximum difference of 15.9%, standard deviations of% differences range from 4.1 to 6.0%). The uncertainty in water fluxes was improved when the standard calibration was adjusted to a stable reference, i.e., the LGR in our study (Fig. 8e-f, maximum difference of 8.9%, standard deviations of% differences ranged from 2.4 to 3.3%), while the best performance was with the piecewise calibration (Fig. 8g-h, maximum difference of 8.2%, standard deviations% of differences ranged from 2.3 to 3.0%). Thus, the uncertainty with the standard calibration is at least 1.4–2.6 × and 1.1–3.7% greater than the minimum uncertainty. As described in the analysis of vertical separation of anemometers/scalar sensors (Section 4.2 below), the water vapor fluxes for the LI-7500 should be interpreted with caution, especially considering it was mounted ~ 2 m lower on the tower and measured eddies that interacted with a slightly different footprint upwind of the scaffold. Without the LI-7500, then for the standard calibration the maximum difference was 11.7% and the standard deviations of% differences ranged from 2.0 to 5.0%, when referenced to the LGR the maximum difference was 6.1% and the standard deviations of% differences ranged from 2.4 to 2.9%, and for the piecewise calibration the maximum difference was 5.1% and the standard deviations of% differences ranged from 1.7 to 2.2%. Without including the LI-7500, the uncertainty with the standard calibration is at least 1.0–2.6 × and 0.0–2.8% greater than the minimum uncertainty. Regardless of considering the LI-7500, around half of the differences in water fluxes can be attributed to uncertainty in the standard calibration.

Finally, the benefit of processing fluxes every 5-minutes instead of a more common 30-minute approach was it allowed more precise fitting of the piecewise calibration and increased the sample size for our analysis; the tradeoff was that a portion of the flux was lost. After removing six outliers (i.e., <1% of the data), 4% of the water flux was lost in the remaining >99% of the data. While the amount of flux loss is of a similar magnitude to the resulting differences in fluxes discussed above, it is important to note that spectral losses due to block averaging tends to occur at low frequencies, while the spectral differences between hygrometers tends to occur at high frequencies (Figs. A1-7). We contend that the flux lost due to 5-minute block averaging was small, affected each hygrometer similarly, and had minimal influence on our results and interpretations.

It can be difficult to synthesize the effect of calibration on H_2O fluxes among other hygrometer intercomparison studies, considering calibration protocols are often ambiguous and most results are confounded by experimental design or the use of different sensors. In one study, Fratini

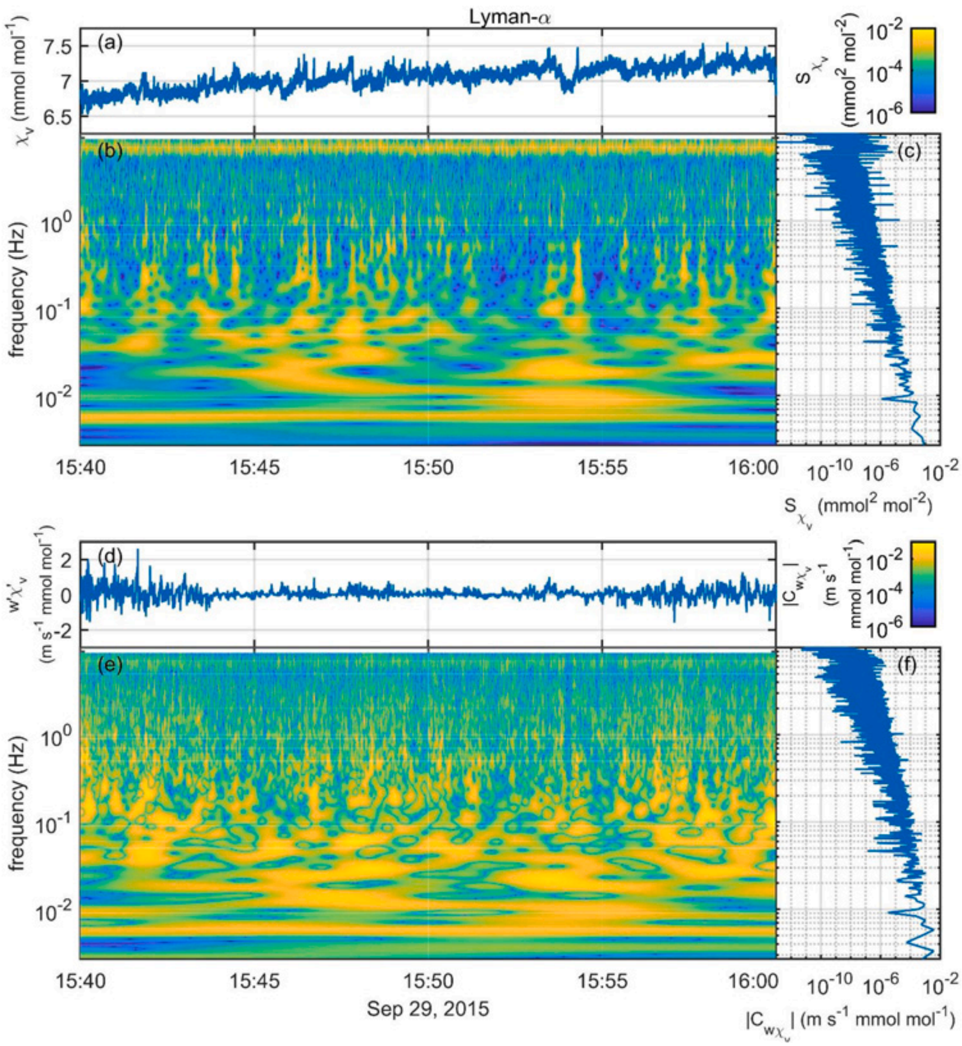

Fig. A1. A half-hour of data from the Lyman-α first generation closed-path hygrometer using the piecewise calibration. (a, d) Timeseries, (b, e) wavelet spectrogram, and (c, f) Fourier decomposition of (a, b, c) vapor concentration and its (d) covariance and (e, f) cospectra with CSAT3 vertical wind velocity. The CSAT3 was upsampled to 20 Hz to match the Lyman-α.

et al. (2014) examined the effect of contamination, etc., on the IRGA water vapor measurement, and demonstrated how this can cause calibration bias; in their field results a bias caused an average 9% difference in water fluxes between a LI-7000 and LI-7200 that increased to a 23% difference as contamination shifted the LI-7200 calibration. Mauder et al. (2006) tested three KH2O and three LI-7500, each receiving a five-point dew point generated lab calibration, and using one of the LI-7500 as a reference; the standard deviation of regression slopes for the KH2O latent heat fluxes was ~ 14% and the standard deviation of regression slopes for the LI-7500 including the reference was ~ 7% (in this study half of the eddy covariance systems included a CSAT3 and the others a USA-1, METEK GmbH, Germany). Mauder et al. (2007) tested six KH2O, using one as a reference, and each receiving pre- and post-deployment lab calibrations; the standard deviation of regression slopes including the reference for the latent heat fluxes was ~ 21% (in this study five different models of sonic anemometers were used). Alfieri et al. (2011) deployed nine replicate eddy covariance sites, each with a LI-7500 that received a zero and span before and after the study and found that for high water fluxes the asymptotic coefficient of variation approached 0.04, i.e., roughly a 4% difference between stations. In this last study calibration uncertainty contributed to at most a 4% difference, which is consistent with our finding that calibration errors added between 1.1–3.7% or 0.0–2.8% uncertainty to our fluxes.

4.2. The effect of hygrometer on measured water fluxes

There was little variation among measured water fluxes attributable to the choice of hygrometer among all sensors that were tested. Considering the piecewise regression mitigates the effect of calibration uncertainty, differences in water fluxes between hygrometers can reveal differences due the underlying technology or sensor design. As described above, comparing fluxes among 2nd and 3rd generation sensors using in the piecewise calibration had at worse a maximum difference of 8.2% and standard deviation of differences of 2.3–3.0%. By including the 1st generation sensors, these difference in fluxes increased to a maximum of 9.4% and a standard deviation of 2.9–3.1% (Fig. 8g-h, all sensors). Many other hygrometer intercomparisons found similar small differences among the same 2nd and 3rd generation sensors: ~ 0% to 6% differences between a LI-7000 and LI-7500 (Ibrom et al., 2007; Kosugi et al., 2007; Wu et al., 2015), < 1% to 4% differences between a LI-6262 (the predecessor of the LI-7000) and LI-7500 (Anthoni et al., 2002; Haslwanter et al., 2009), < 1% difference between a LI-7500 and LI-7200 (Burba et al., 2010), and a 2% difference between a LI-7200 and EC155 (Novick et al., 2013). In a few studies the differences were larger than we observed: a 7% to 12% difference, depending on the tube attenuation correction, between a LI-7000 and LI-7500 (Fratini et al., 2012) and a −12% to +5% difference, depending on u* filtering, between two LI-7500 s and a EC155 (Novick et al., 2013). Kutikoff et al. (2021) found 0% and

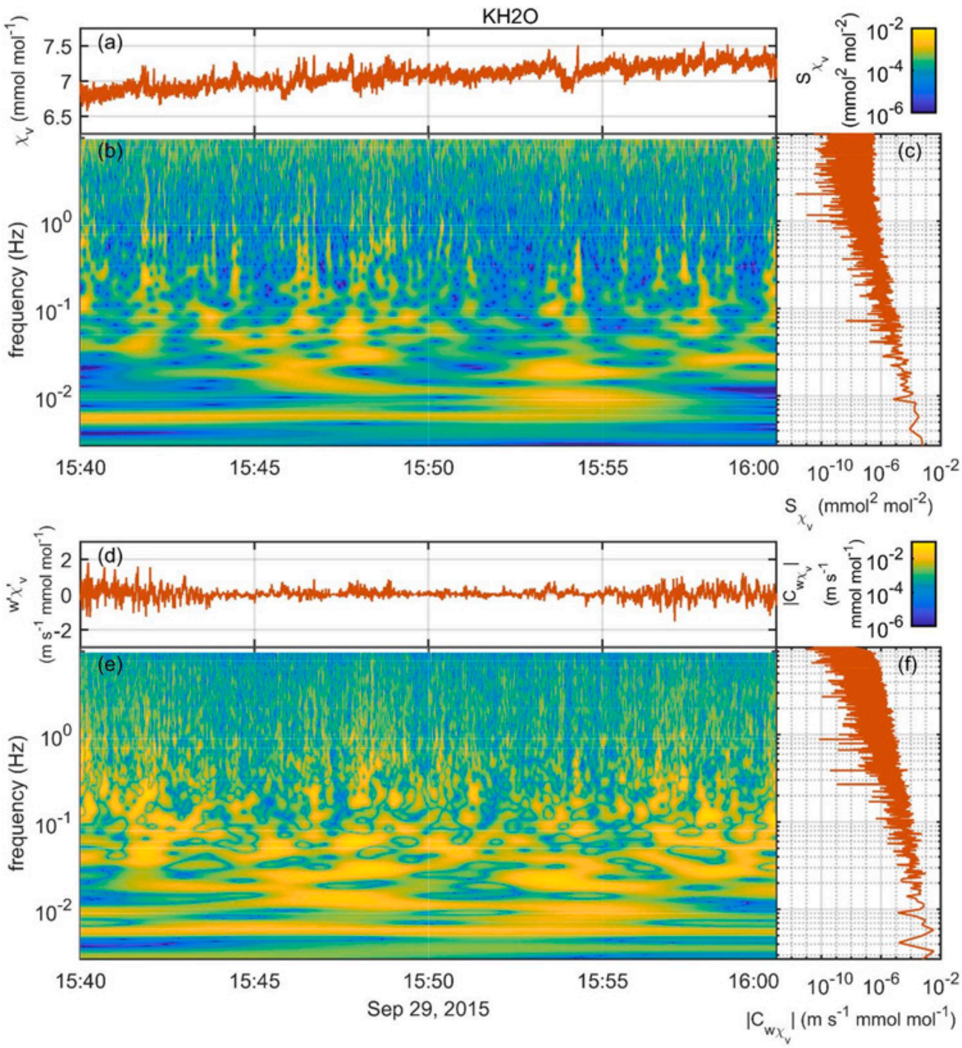

Fig. A2. A half-hour of data from the KH2O first generation open-path hygrometer using the piecewise calibration. (a, d) Timeseries, (b, e) wavelet spectrogram, and (c, f) Fourier decomposition of (a, b, c) vapor concentration and its (d) covariance and (e, f) cospectra with CSAT3 vertical wind velocity. The CSAT3 was upsampled to 20 Hz to match the KH2O.

12% differences in latent heat fluxes among three generations of the LI-7500, with the oldest sensor producing the largest measurements. In a large intercomparison, Polonik et al. (2019) investigated a LI-7500A, LI-7200, CPEC (i.e., an EC155), an IRGASON (i.e., an EC150), and a Picarro cavity ring-down sensor (i.e., similar to our LGR) and found an average of 7% difference between H_2O fluxes when omitting their Picarro and some of their LI-7500A data and using the Massman (2000) spectral corrections, i.e., a similar methodology to ours. Their differences are $\sim 3\times$ greater than our results, though this could be due to calibration uncertainty (e.g., they mention sensor drift or poor calibration as a potential cause of their questionable LI-7500A results), their fluxes are calculated with a combination of three sonic anemometers, of which their CSAT3s were not transducer shadow corrected, and, as they discuss, their measurements were at 4 m above short vegetation where fluxes are more dependent on high-frequency eddies that require larger and more uncertain spectral corrections than at our 22.65–25.2 m heights above a forest.

The biggest source of uncertainty in the processing of our data was the O_2 correction for the Lyman-α and krypton hygrometers (van Djik et al., 2003), especially for the KH2O which had the highest corrected water fluxes (Fig. 8g-h). When we processed the fluxes without this correction (Fig. 8i-j), the KH2O dramatically shifted to produce among smallest water fluxes (a relative shift of $\sim$ 9%) and the Lyman-α fluxes decreased by $\sim$ 3%. While the pathlengths of the KH2O and Lyman-α were relatively small, the 11.1 mm of the KH2O was more of an issue than the 2.4 mm of the Lyman-α (Beaton and Spowart, 2012). Whether the KH2O tends to measure larger or smaller fluxes is difficult to assess from the literature. Mauder et al. (2006) tested three KH2O and three LI-7500, and the average regression slope comparing latent heat fluxes was $\sim$ 17% higher for the KH2O. Mauder et al. tested six KH2O and one LI-7500, and the average regression slope comparing latent heat fluxes was $\sim$ 24% higher for the LI-7500. Martínez-Cob and Suvočarev (2015) compared a KH2O and LI-7500 and found $<$ 2% difference in water fluxes. Judd et al. (1993) compared a KH2O and LI-6262 and found no difference in their fluxes, though the slope of the relationship drifted by $\sim$ 36% over the course of their study; uncertainty of this magnitude is consistent with our findings that the KH2O had severe calibration drift and a large and uncertain O_2 correction. Lee et al. (1994) compared the same instruments and found 7 and 11% differences at separate field sites, though they attributed the difference more to the tube attenuation in the LI-6262. We investigated the role of the tube-attenuation correction for closed-path analyzers as another source of uncertainty in water fluxes (Fig. 8k-l). The EC155 had the lowest flow rate, which manifests in a depression of high frequencies in the power spectra (Fig. A6c). The EC155 had the largest tube attenuation correction, and without it the fluxes decreased relative to the other hygrometers by $\sim$

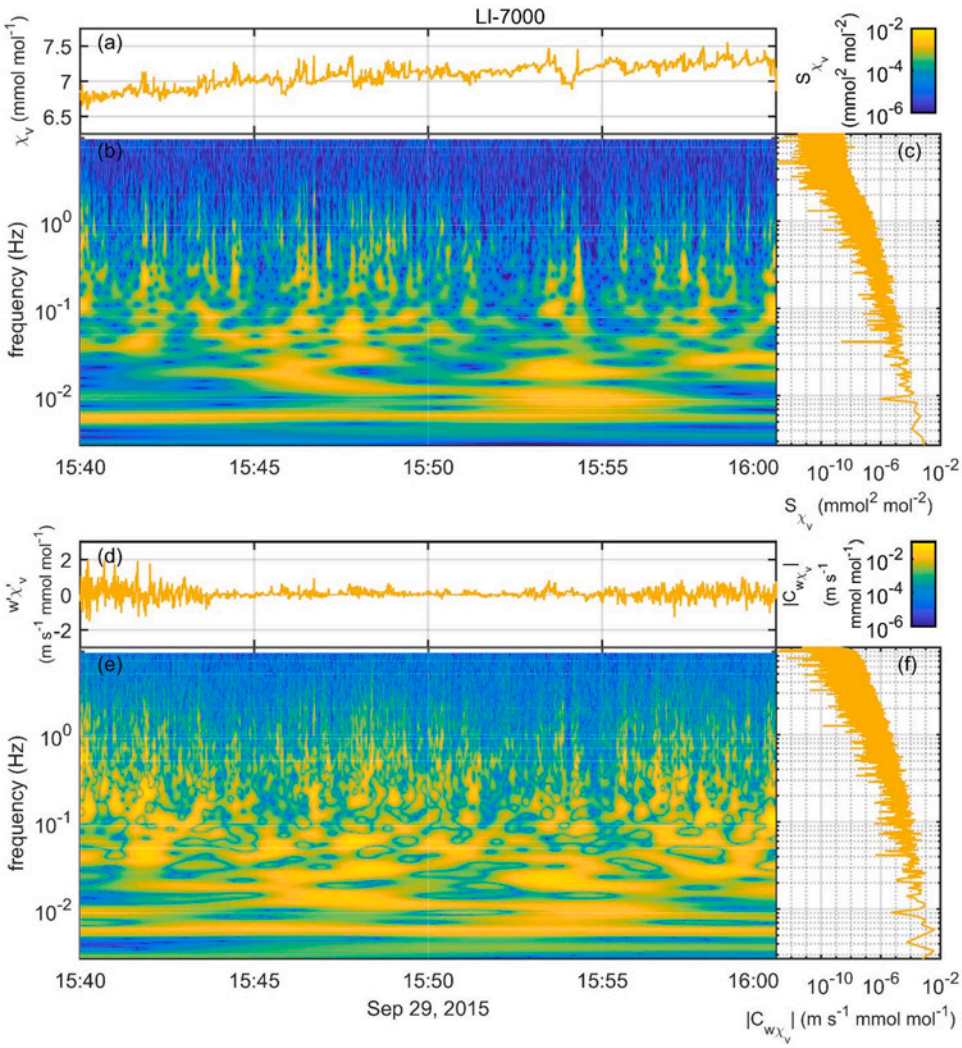

Fig. A3. A half-hour of data from the LI-7000 second generation closed-path hygrometer using the piecewise calibration. (a, d) Timeseries, (b, e) wavelet spectrogram, and (c, f) Fourier decomposition of (a, b, c) vapor concentration and its (d) covariance and (e, f) cospectra with CSAT3 vertical wind velocity. The CSAT3 was upsampled to 20 Hz to match the LI-7000.

1.4% using the CSAT3 (Fig. 8k) and ~ 0.4% using the ATI 3Vx-probe (Fig. 8l). A potential reason for the smaller effect with the ATI 3Vx-probe is that tube attenuation is not the only major source of high frequency loss, and our spectral corrections already correct much of the high frequency attenuation due to the horizontal separation between sensors. Spectral corrections for tube attenuation can be a substantial source of H_2O flux uncertainty in eddy covariance systems with long tubes (Ibrom et al., 2007; Runkle et al., 2012) or with measurements close to the surface over short vegetation (Fratini et al., 2012; Polonik et al., 2019). Together, these studies found problems with tube attenuation in the LI-7000, LI-7200, and EC155. In our study only the EC155 experienced a significant effect of the tube attenuation correction. This sensor, along with the LI-7200, are often referred to as "enclosed path" instead of closed path because they are designed to be deployed outside and near the sonic anemometer to reduce the effects of tube attenuation. In our study all closed path tubes were unheated but relatively short; heating intake tubes can help avoid errors associated with tube attenuation (Ibrom et al., 2007). If it were assumed the main differences in fluxes between open- and closed-path hygrometers were due to insufficiently corrected tube attenuation, this error can be estimated relative to the EC150 (i.e., the only open-path hygrometer among the co-located 2nd and 3rd generation sensors): 0.3% less water flux for the LI-7000, 2.5% less for the LI-7200, and 3.5% less for the EC155 (Fig. 8g). Of these potential errors, the only one that we could specifically explain was a 1–2% error in the EC155 due to tube attenuation at flows of 7 LPM, though this is not an uncommon setting (Novick et al., 2013; Polonik et al., 2019; Wu et al., 2015). Otherwise, without a clear reason to attribute the rest of the 0.3–3.5% differences between the open- and closed-path hygrometers to tube attenuation, there little evidence that it was a major problem in our experimental setup with short tube lines, high flow rates, and fluxes with a somewhat lower cospectral peak frequency due to the scaffold height. Overall, the O_2 correction for the KH2O was our most significant source of uncertainty in water fluxes attributable to a fundamental difference in hygrometer design.

One complication with interpreting our results is that the LI-7500 and ATI 3Vx-probe were not co-located with the other equipment during this study, rather they were dedicated to US-GLE eddy-covariance system (Fig. 1). The benefit of including the LI-7500 in this study is that it is perhaps the most common IRGA used for eddy covariance studies. The benefit of including the ATI 3Vx-probe is that until the CSAT3 was installed during the last week of our study, it was the only sonic anemometer available to calculated water vapor fluxes. To improve comparisons between ATI and CSAT3 sonic anemometers (Frank et al., 2013; Frank et al., 2016a, Kochendorfer et al., 2012) we applied the transducer shadowing algorithm of Wyngaard and Zhang (1985) and described in Horst et al. (2015). To improve comparisons between

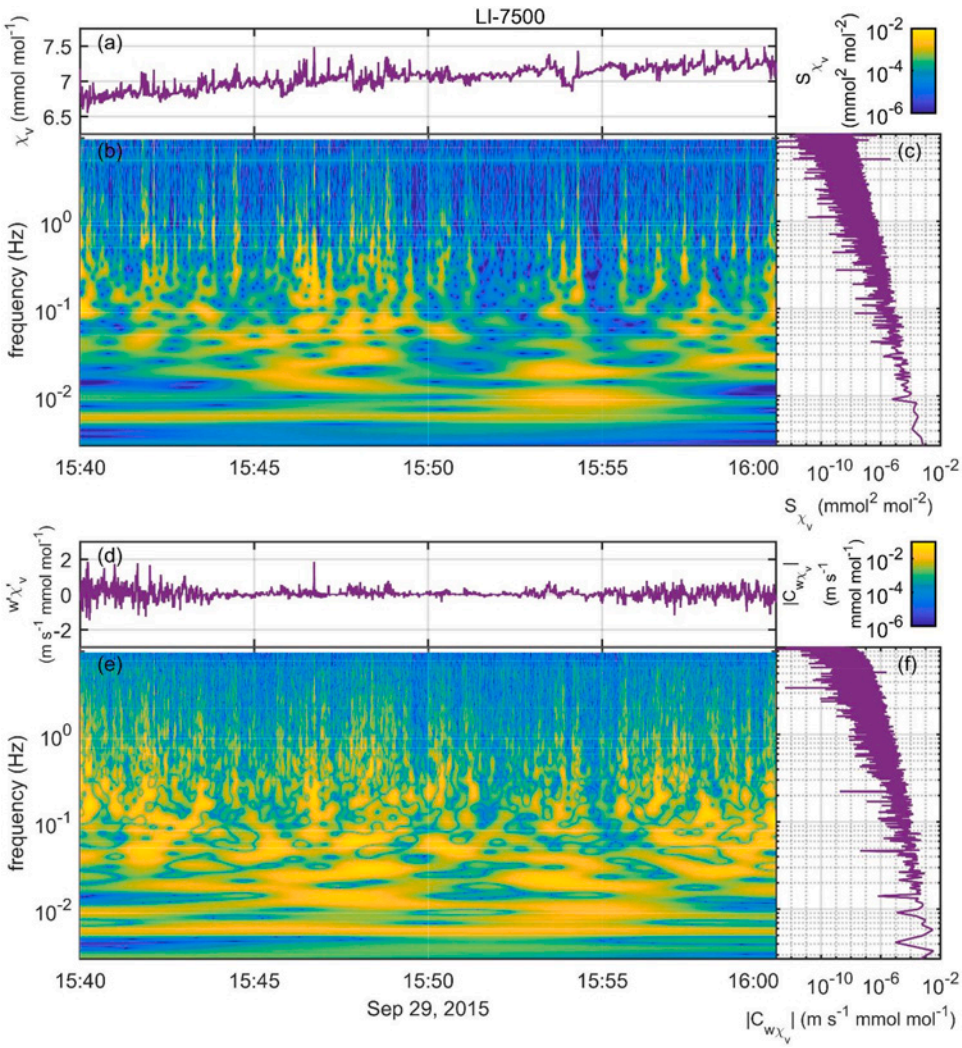

Fig. A4. A half-hour of data from the LI-7500 second generation open-path hygrometer using the piecewise calibration. (a, d) Timeseries, (b, e) wavelet spectrogram, and (c, f) Fourier decomposition of (a, b, c) vapor concentration and its (d) covariance and (e, f) cospectra with CSAT3 vertical wind velocity. *The LI-7500 was located ~ 2.5 m below the CSAT3 and the other hygrometers (Fig. 1). The CSAT3 was upsampled to 20 Hz to match the LI-7500.

horizontally separated sensors we applied the spectral correction for lateral separation described in Massman (2000), which resulted median spectral corrections of +8% for the LI-7500 and between +9 to +13% for the others (i.e., the LI-7500 was closer to the CSAT3 than the other hygrometers were to the ATI 3Vx-probe). We did not correct for the vertical separation between anemometer and hygrometer. Kristensen et al. (1997) describe that flux loss should be −1.1% at our heights when the scalar sensor is mounted below the anemometer (i.e., the LI-7500 sensor below the CSAT3) whereas when the scalar sensor is above the anemometer (i.e., all other hygrometers above the ATI 3Vx-probe) the flux loss would be −10.1% (Table 3). Thus, we would not expect water fluxes between the LI-7500 and CSAT3 to be significantly reduced, and in our analysis these fluxes were only ~ 2% less than the range of the other 2nd and 3rd generation sensors (Fig. 8g). Based on Kristensen et al. (1997) we would have expected substantial flux loss when using the ATI anemometer for all but the LI-7500, though our results did not show this and, in fact, the water vapor flux from the LI-7500/ATI 3Vx-probe was less than with the other hygrometers. We further examined vertical separation by focusing on sensible heat flux and calculating covariances between the vertical wind velocity of one sonic anemometer and the virtual temperature from the other. In both cases of vertical separation in the sensible heat flux measurement, the resulting fluxes were about 7% larger than predictions based on Kristensen et al. (1997) (Table 3). Though, like predictions based on Kristensen et al. (1997) where the scalar below the sonic should produce 9% higher fluxes than when it is above, our sensible heat flux with the scalar below was 6% higher than with the scalar above (Table 3). One possibility is our corrections for horizontal separation were too large. Had the corrections between the ATI 3Vx-probe and the CSAT3 been 4.6% and 2.7% instead of 9.8% and 11.2% the results would have been consistent with Kristensen et al. (1997). Another possibility is that the lower mounted ATI 3Vx-probe measures a slightly different and smaller footprint of the subalpine forest that generates different amounts of sensible heat flux; to adequately quantify this would be beyond the scope of this study. If smaller horizontal corrections had been applied to water vapor fluxes and the 1.1% and 10.1% amount of flux loss predicted by Kristensen et al. (1997) added back in, the resulting LI-7500/CSAT3 fluxes would be ~ 4% smaller (i.e., moved to the left Fig. 8g) and all other fluxes with the ATI 3Vx-probe would be ~ 1–2% larger (i.e., moved to the right in Fig. 8h). Though, in both circumstances this would widen the gap between the LI-7500 fluxes and the other hygrometers. Therefore, we interpret the LI-7500 fluxes with caution. While it is encouraging that the LI-7500 fluxes were similar to the other hygrometers, it is possible that they would have been different had the LI-7500 been co-located with the other sensors.

Overall, we interpret that after controlling for uncertainties in

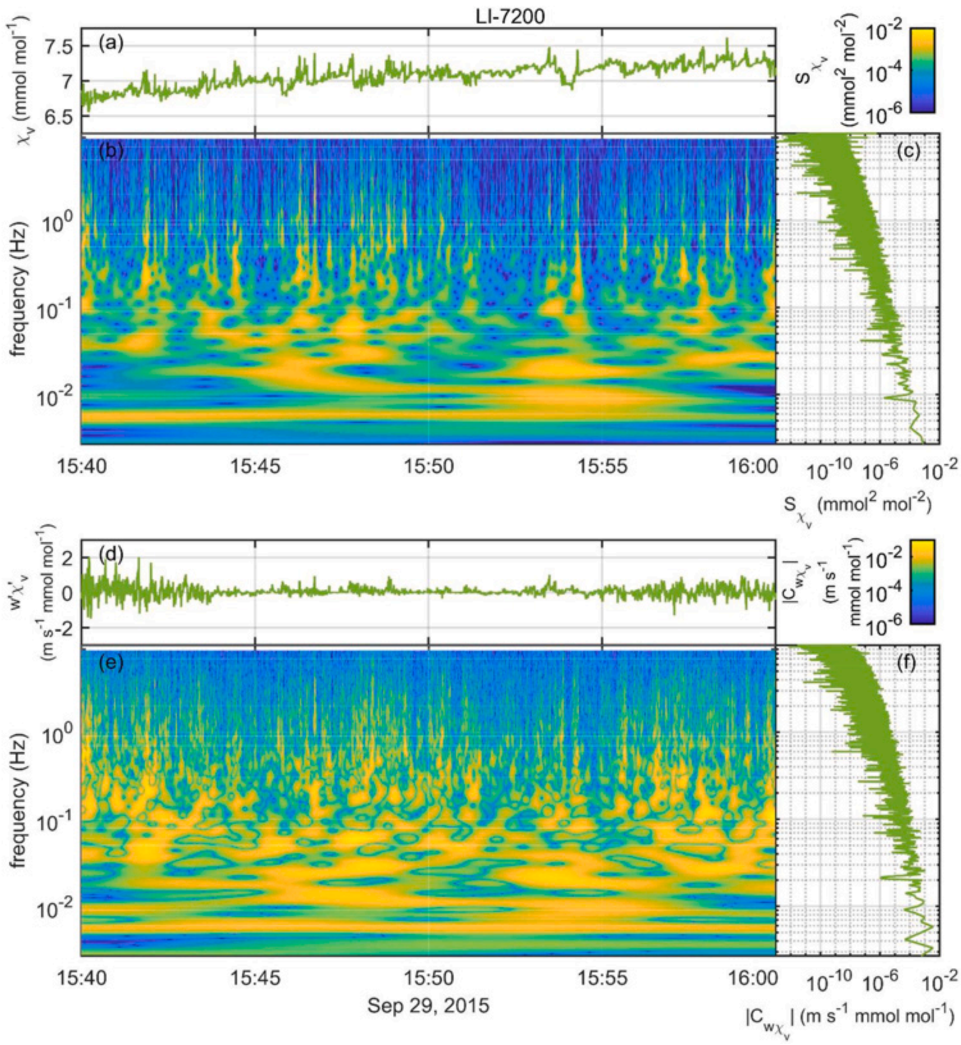

Fig. A5. A half-hour of data from the LI-7200 third generation closed-path hygrometer using the standard calibration. (a, d) Timeseries, (b, e) wavelet spectrogram, and (c, f) Fourier decomposition of (a, b, c) vapor concentration and its (d) covariance and (e, f) cospectra with CSAT3 vertical wind velocity. The LI-7200 was the reference to which all other hygrometers derived their piecewise calibration. The CSAT3 was upsampled to 20 Hz to match the LI-7200.

calibrations, there is very little difference in the accuracy of water fluxes measured with 2nd and 3rd generation hygrometers. And considering that both 1st generation hygrometers suffered from dramatic calibration drifts, and in the case of the KH2O, required a substantial oxygen correction, water fluxes measured with these hygrometers can be made accurate. It is remarkable that most of these hygrometers have unique signal issues, which are visible in their Fourier transforms and wavelet spectrograms as shown in the appendix, that ultimately have little bearing on the water vapor flux. In all circumstances, the covariance with vertical wind velocity tends to filter out most of these problems, such that the cospectra for each hygrometer tend to look similar (Figs. A1–7 panels e-f). For example, the Lyman-α, which has an elevated noise floor and numerous high-frequency noise peaks, and the LI-7000, which has a dampened response, have dramatically different power spectra as shown in their wavelet spectrograms above ~ 0.4 Hz (i.e., different colorations in the upper halves of Figs. A1b vs A3b), yet the corresponding plots for their cospectra are nearly identical at all frequencies (Figs. A1e vs. A3e) as both capture the same covarying structures related to water vapor exchange.

It is important to consider that these results are from 22.65 to 25.2 m above a canopy that is ~ 7–18 m tall. Under these conditions the total flux is dominated by eddies circulating at frequencies that are generally lower than where the various flaws occur among these hygrometers. It would be likely that a shorter tower over shorter terrain would experience a larger contribution from higher frequency eddies and would likely have a larger discrepancy in water vapor fluxes using different hygrometers. As it is, beyond the issues of O_2 correction, mainly for the krypton hygrometer, and some tube attenuation due to lower flow rates, we found minimal impact of sensor technology on the variation in water vapor fluxes.

4.3. The effect of hygrometer on energy balance closure

In general, our results regarding the energy balance closure fall within four categories: closure is most variable amongst hygrometers using the standard calibration, differences are negligible using the piecewise calibration implying limited effect of hygrometer technology, referencing calibrations to a first principles standard tended to increase the closure, and the difference between using an orthogonal versus a non-orthogonal anemometer corresponded to better closure. While the sample size for testing energy balance for all seven hygrometers was small, the largest discrepancy in the closure ratio occurred when using the standard calibration, which was 0.022 between the EC150 and LI-7000; the standard deviation of closure among all hygrometers was 0.009. For the piecewise calibration, the largest discrepancy was 0.021, and the standard deviation of closure among all hygrometers 0.008;

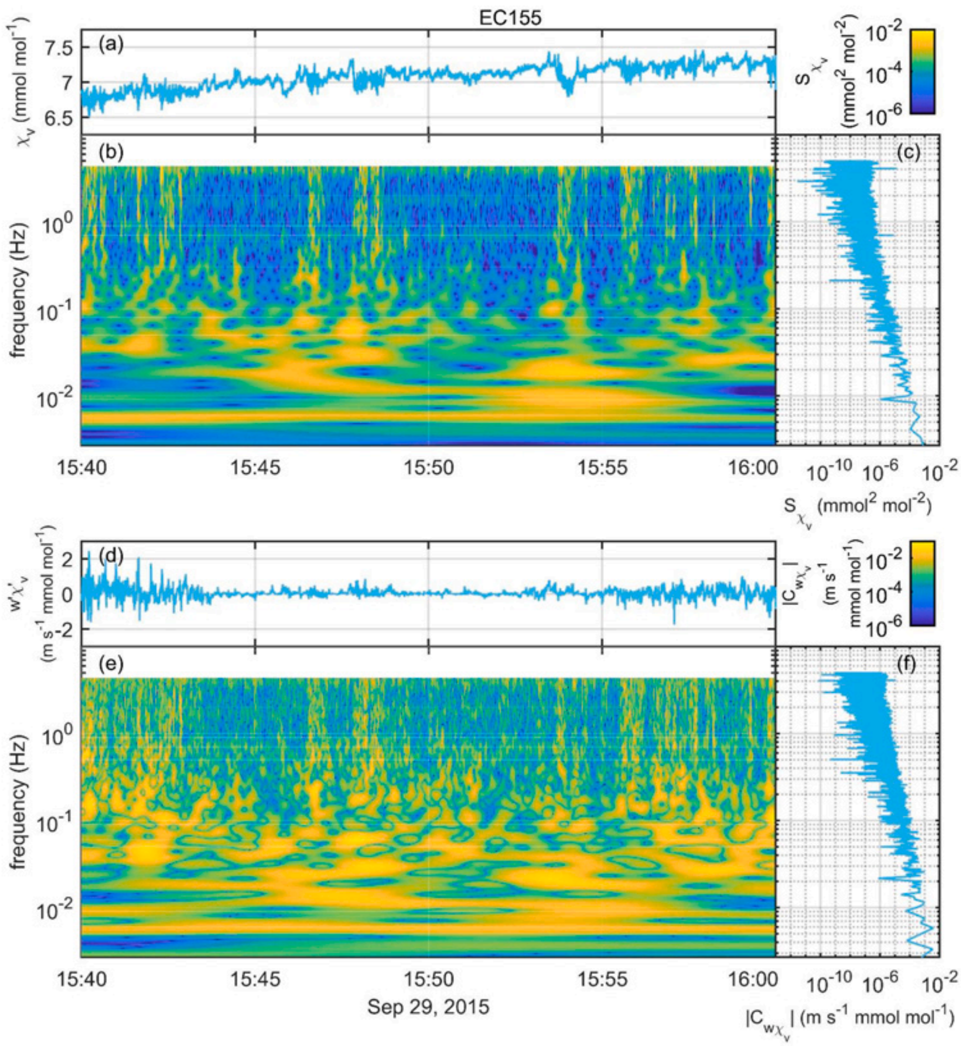

Fig. A6. A half-hour of data from the EC155 third generation closed-path hygrometer using the piecewise calibration. (a, d) Timeseries, (b, e) wavelet spectrogram, and (c, f) Fourier decomposition of (a, b, c) vapor concentration and its (d) covariance and (e, f) cospectra with CSAT3 vertical wind velocity. The EC155 and CSAT3 were sampled at 10 Hz.

both are probably negligible uncertainties for most eddy covariance data users. In this study we referenced the piecewise calibration to the standard calibrated LI-7200, which was among the most underestimated when compared to the LGR (Table 2), i.e., the LI-7200 might be underestimated by ~ 11% and thus all the piecewise calibrated hygrometers may also be underestimated by ~ 11%. Accounting for this raised the energy balance closure by 0.023. And finally, when energy balance was calculated using the ATI 3Vx-probe instead of the CSAT3 the closure ratio increased by 0.048. This is not surprising, as Frank et al. (2016a) observed that even a transducer shadowing corrected CSAT3 can still underestimate fluxes by ~ 5% when compared to an ATI anemometer.

Considering the energy balance closure ratio across flux sites is highly variable, but often reported as a ~ 20% bias towards underestimates in the turbulent fluxes (Leuning et al., 2012; Stoy et al., 2013; Wilson et al., 2002), there are several insights from our study. First, regardless of the method of calibration, there was minimal difference in energy balance closure between hygrometers; the standard deviation among closures ratios was ~ 1% with a worst-case difference of ~ 2%. Our results could also have been biased, as demonstrated by the ~ 2.5% lower closure if we consider that the LI-7200 was underestimated relative to the LGR. As four of the fast-response hygrometers were underestimated relative to the LGR (Table 2), this might be a systematic problem in implementing the standard calibration. The small effect of hygrometer differences on energy balance has between observed elsewhere: Wu et al. (2015) measured a 3% difference in closure using a LI-7000 versus a LI-7500 while Haslwanter et al. (2009) found a 2% difference with a LI-7500 versus LI-6262. It is worth putting into context, though, that the largest contribution to the energy balance uncertainty we found was the 5% increase corresponding to calculating fluxes with the ATI 3Vx-probe versus the CSAT3. While this study had to deal with the vertical separation of sensors and was less rigorous than others conducted between the CSAT3 vs ATI anemometers (Frank et al., 2013; Frank et al., 2016a) it is worth acknowledging that the greater uncertainty in closure, even after the transducer shadowing was applied to the CSAT3 (Horst et al., 2015), corresponds more with the choice in sonic anemometer than any of the calibration routines in the fast-response hygrometers.

The largest closure ratio in this study was obtained using the piecewise calibration, referencing everything the first principles based LGR, and using the ATI 3Vx-probe. In this situation the closure was 88%, which would be considered better than typically found across flux networks. In the studies mentioned above, Wu et al. (2015) and Haslwanter et al. (2009) measured energy balance closures between 73 and 82%. Considering that our 5-minute fluxes resulted in 2% lower H and 4–8% less LE compared to 30-minute fluxes, it is likely that our energy balance

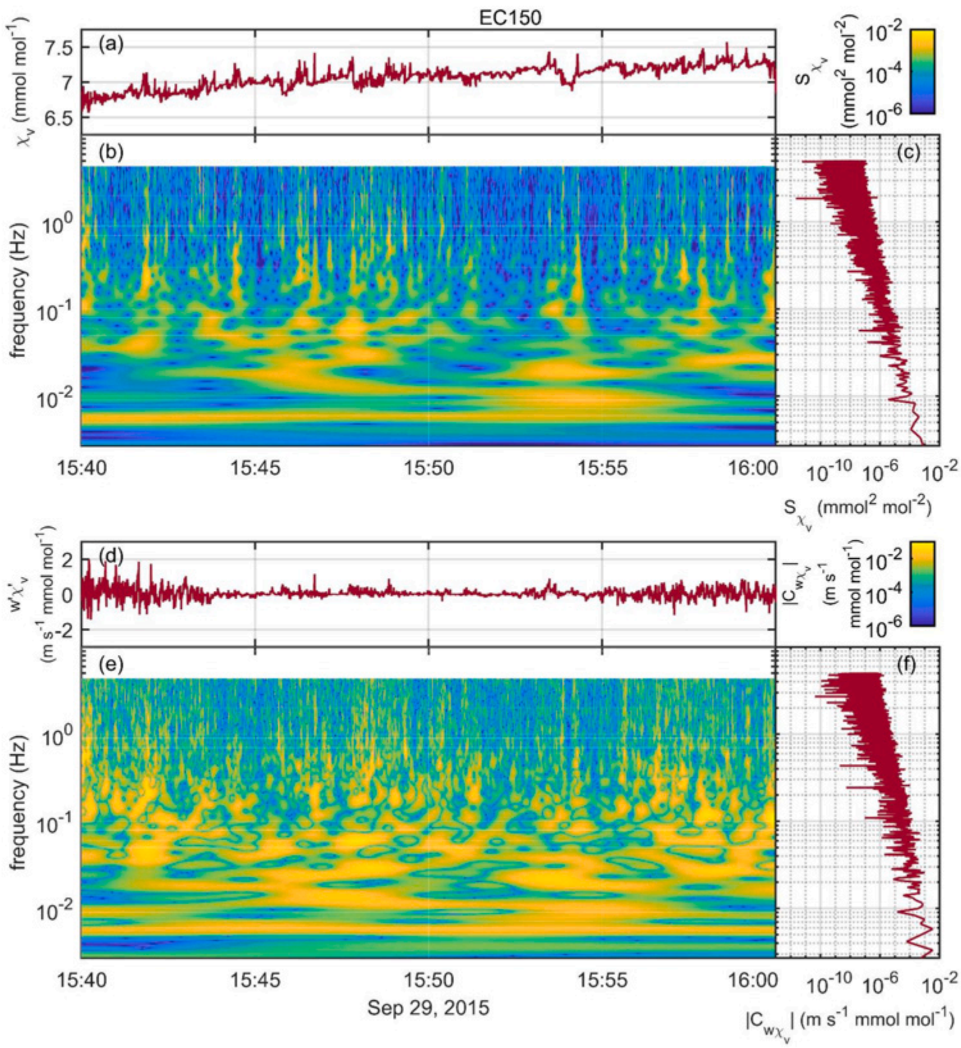

Fig. A7. A half-hour of data from the EC150 third generation open-path hygrometer using the piecewise calibration. (a, d) Timeseries, (b, e) wavelet spectrogram, and (c, f) Fourier decomposition of (a, b, c) vapor concentration and its (d) covariance and (e, f) cospectra with CSAT3 vertical wind velocity. The EC155 and CSAT3 were sampled at 10 Hz.

closure would be slightly closer to unity using a 30-minute approach. Though, using longer flux calculation periods in our study would have resulted in smaller sample sizes which were already problematic (i.e., $n = 175$ in Fig. 9a-c, $n = 236$ in Fig. S4a-c). The ability to generalize our study is limited considering that GLEES is a relatively arid site, and during the late growing season when our data was collected the Bowen ratio was ~ 3. We anticipate that wetter sites will have energy balance closures that are more sensitive to differences in water fluxes.

5. Conclusion

We based our approach on a relative framework, where multiple instruments were compared to each other, but not necessarily to an absolute standard. This is because, despite our intentions, our study was hampered in measuring an accurate and true standard for humidity. For all hygrometers we conducted at least one in situ standard calibration with a dew point generator, and except for noting the discrepancies regarding the zero-air point (Table 1 and Fig. 2), each procedure would have been considered a success. But the context provided by the other hygrometers demonstrated that relying on the standard calibration to be an absolute reference can be troublesome. While clearly the problems with the Lyman-α and KH2O (Fig. 5a) went far beyond drifts in a few dew point generator calibrations (none of which are presented in this paper) the issues with the other hygrometers were more concerning (Table 1, Fig. 8a-d). While our standard calibration procedure could have been enhanced by conducting them more frequently (i.e., daily, hourly), performing experiments such as the effect of flow rate on the generated output humidity, or testing the accuracy of the dew point generator against another standard such as the LGR, etc., none of these would occur at a typical field site. Rather, the variability we found using the standard calibration, both in repeated calibrations of the same hygrometer and in the relative differences between hygrometers, probably occurs at most field sites, where, without multiple sensors, it would be difficult to detect this uncertainty. It may be erroneous to assume a hygrometer that has been calibrated against a dew point generator retains the accuracy of an absolute standard. The intended absolute standard for this study was a chilled mirror hygrometer, which is designed to measure humidity using first principles. We deployed two of these sensors, but instabilities in their electronics caused calibration drifts between them of a similar amount to the drifts we found in the fast-response hygrometers; we do not present data from these sensors in this paper. The closest we came to an absolute standard was the LGR cavity ring-down isotope analyzer which is based on first principles and theoretically does not require intermittent calibration. This sensor was not an ideal standard for this study because of its slow response (~ 90 s time constant) and because it was deployed as part of a profiling study which intermittently sampled air from up and down the scaffold. Newer cavity ring-down sensors have faster response times and would be more suitable for such an application (Polonik et al., 2019). Yet, in our study the LGR was useful to identify the LI-7200 as having the least amount of calibration drift over time and to demonstrate that most of the standard calibrations were underestimated (Table 2). As it was, the most stable and directly comparable sensor was the slow-response capacitance humidity sensor (Met-One 083D) used for the basic meteorology of the site. While the full-scale accuracy of this sensor is questionable, especially for very low and very high humidity, when compared to the first principle based LGR, it was reasonably accurate and stable (Table 2).

Considering our most favorable results involved the piecewise calibration, we consider whether it could be implemented at a typical flux site. Though our study focuses on duplicate fast-response hygrometers, of which a typical site would have just one, most sites also have a slow-response (often capacitance) humidity sensor. These sensors can retain accurate and stable calibrations over time; in our study only the LI-7200 had comparable long-term stability to our capacitance sensor (Table 2). To successfully compare fast- and slow-response sensors, consideration must be given to the timescales at which their measurements should be similar. For this, a quantile approach would likely characterize fast- and slow-response sensors differently and, thus, is likely inappropriate. Therefore, we investigated the use of medians (or interchangeably means) in the piecewise calibration. The quality of the piecewise calibration using medians was slightly worse than when using quantiles. Objectively, this was quantifiable through various statistical parameters: the standard deviation of the% differences among the 2nd and 3rd generation sensors (2.8 and 3.4%) was slightly higher than when using quantiles (2.3 and 3.0%), the average σ, i.e. the standard deviation of the prediction, was 24% larger, and the standard deviations of the posterior distributions for the% differences among fluxes were 15–47% larger (Figure S5g vs. 8 g). Subjectively, this was apparent at the start/end of each block when it was obvious that some predicted fluxes were implausible (hence, the additional QAQC step described at the end of Section 2.2). Unlike quantiles which can reveal an accurate calibration between fast-response sensors within a 5-minute period, the median approach requires multiple 5-minute periods to establish a trend across, and even then, the range of data that is considered is often less than spanned by the quantiles. Regardless, we found the median approach to the piecewise calibration outperformed the standard calibration by reducing the% differences among hygrometers by 0.4–0.5 ×.

For any application of the piecewise algorithm, the choice of threshold is critical to the overall performance, yet determining an optimal value is somewhat subjective. In general, the following guidelines should apply. Starting with a high threshold and progressing towards lower values, (1) the number of blocks will increase to a maximum before eventually decreasing, (2) the amount of data included in blocks will decrease, (3) the RMSE between the piecewise calibrated data and the reference data will decrease. The rates of decrease for (2) and (3) are not the same, as (2) tends to decline slowly until a certain threshold is reached after which it drops rapidly while (3) decreases somewhat linearly with the threshold. Strategies to determine an optimal performance could include a-priori choosing a target amount of data included, RMSE between sensors, or determining the point at which the amount of data included drops rapidly. A further consideration is that the piecewise coefficients can be interpolated to include all the data. This would be pragmatic for a typical field site where it is desirable to obtain flux calculations from a sensor that otherwise passes all data quality flags. Because (4) the RMSE of the data derived from linearly interpolated piecewise coefficients is always higher than from data within the blocks, (5) as the threshold is lowered there is a local minimum in the RMSE for all data (i.e., both included in blocks and interpolated between blocks). For a typical field site, finding this local minimum in RMSE might be the most desirable strategy, as it represents the lowest overall uncertainty in calibrating the complete dataset. Often these strategies yield different results. For the LI-7000, strategizing around (2) yields a threshold around 0.00008 mmol mol^{-1} where the RMSE of data within the blocks is 0.0184 mmol mol^{-1} and the amount of data begins to drop quickly from 92%, while the local minimum in (5) corresponds to a threshold of 0.0002 mmol mol^{-1} where the RMSE of data within the blocks is 0.0411 mmol mol^{-1} and the amount of data used is 96%. For the KH2O, strategizing around (2) yields a threshold around 0.0002 g m^{-3}, a within block RMSE of 0.0143 g m^{-3}, and uses 79% of the data while (5) corresponds to a threshold of 0.00028 g m^{-3}, a within block RMSE of 0.0181 g m^{-3}, and uses 88% of the data. As our study focused on removing sources of uncertainty, we utilized only data identified within the piecewise blocks, and hence focused on strategizing around (2). Hence, our thresholds are 20–70% smaller than an approach that is optimized around (5) to use all of the data via interpolation of the calibration coefficients. Finally, we found few differences among these strategies when using performing a piecewise calibration using medians, such as would be done with a slow-response hygrometer, beyond the identification of fewer blocks and the necessity to use lower thresholds to achieve similar quality of fit. Hence, in our analysis we chose thresholds that were 25–50% less when using medians instead of quantiles.

Based on this study, we recommend the following improvements for more accurate water vapor flux measurements. First, an accurate and stable humidity sensor is important. This need not be a fast-response eddy covariance sensor, as it likely requires a sensor that is more stable over time. In our study the most stable hygrometers were the LGR, LI-7200, and Met-One 083D (Table 2). While the first principles, no-calibration required operation of the LGR was appreciated, we acknowledge that its size, shelter requirements, and power consumption, not to mention expense, make it impractical for most applications. We encourage the micrometeorology community to continue to develop programs that provide accurate calibration of humidity sensors, much in the same way that is currently done with CO_2 calibration gas tanks or PPFD sensors (https://ameriflux.lbl.gov/tech/support-services/).

Second, the micrometeorology community should continue to develop common practices for calibrating hygrometers. In our study, it was difficult to determine which calibrations were more successful between open-path and closed-path sensors. The lack of repeatability of the standard calibration of the LI-7200 and LI-7000 calibrations just a week apart was troubling, as it either implicated drifts in the sensors or failures in dew-point generator methodologies. The inability to generate dew points that occur naturally was also troubling, as was the change in calibration slopes and the quality of fit when adding a zero-air point. Fratini et al. (2014) investigated the theory of operation of IRGAs and analyzed the effect of contamination and other effects on calibration bias. They concluded that water vapor field calibrations were error prone because it is difficult to control the microclimate surrounding the sensors, e.g., temperature, ventilation, radiation loads, etc., yet they recommend checking the zero-air point periodically. We cannot disagree with their conclusion considering the large uncertainty in our fluxes based on our field calibrations.

Third, the community should consider whether post-hoc calibrations versus an ambient standard would be preferred over a more conventional calibration to provide the most accurate water vapor fluxes. As shown in this study, even the most stable hygrometers appear to drift over time, and even the best-intentioned dew-point generator calibrations can cause more uncertainty than they solve. Sensors are also prone to abrupt shifts such as in Fig. 4a, Fratini et al. (2014), and Kutikoff et al. (2021). Loescher et al. (2009) demonstrated that hygrometer calibrations can be updated on the order of seconds/minutes/hours, while Frank et al. (2014) did something similar in performing an auto-regression over days/months/years between the fast-response hygrometer and the slow-response capacitance humidity sensor. In both cases, the ability to post-hoc calibrate the sensor is limited by the natural variation of the atmosphere. Fratini et al. (2014) described a correction procedure that combined calibration-check data and differences between IRGAs (they used a LI-7000 as a reference) to create a time series of absorptance biases from which to correct drifts due to contamination, etc. A potential spin-off of the present paper could be the use of a piecewise calibration that identifies blocks of time to compare a fast-response hygrometer against a standard. Possible enhancements could include allowing blocks to incorporate a change in calibration over time such that a piecewise approach could differentiate between abrupt calibration shifts versus slower drifts. If appropriate, the use of quantiles and 5-minute processing provides more samples and wider variation of naturally occurring humidity to base a post-hoc calibration upon. We also comment that such an approach should probably be done with water vapor mixing ratio, e.g., mole fraction in moist air (Fratini et al., 2014), as this quantity should be conserved regardless of the type of sensor. This means for a simple capacitance humidity sensor it must be accompanied by an accurate temperature and pressure measurement.

In summary, we are encouraged that there was minimal evidence to support that water vapor flux measurements are meaningfully different among common hygrometers in use today, as well as historically important sensors. We were discouraged at the difficulty in obtaining a stable and accurate humidity reference, as well as the apparent calibration drift among most of the fast-response hygrometers. After constructing a piecewise calibration to adjust all hygrometers to have similar measurements, the 2nd and 3rd generation fast-response hygrometers yielded water vapor fluxes that were $\sim$ 2–3% different from each other, while the 1st generation sensors were similar but slightly more uncertain. Even with the most uncertain calibrations, the impact of the hygrometer on the energy balance closure was small, and with the piecewise calibration it was negligible; the choice of sonic anemometer had a larger impact. As new generations of fast-response hygrometers become more reliable, a focus should be on improving methods for accurate calibration, such as referencing to a known standard to correct absorptance bias (Fratini et al., 2014) or further developing a piecewise calibration to correct sensor drift.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

We thank Stuart Beaton of the National Center of Atmospheric Research for providing the Lyman-α hygrometer, Brent Ewers and David Williams of the University of Wyoming for providing the LI-7000, LI-7200, and LGR, and Ivan Bogoev of Campbell Scientific, Inc. for providing the EC155, EC150, CSAT3, and accompanying data acquisition system. We thank Thomas Foken and Andreas Ibrom for their thoughtful reviews that improved this manuscript. This study was funded by the USDA Forest Service.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.agrformet.2023.109437.

Appendix A. Timeseries, Fourier, and wavelet visualization of hygrometer signals

To provide insight into the performance of the fast-response hygrometers, the time series, Fourier transforms, and wavelet spectrograms for the hygrometer signals and their covariance/cospectra with vertical wind velocity are provided in Figs. A1-A7. A single representative twenty-minute period was chosen with moderately strong water vapor fluxes and while all hygrometers had a successful piecewise calibration. The Fourier transform of the hygrometer signal is presented as the power spectra. In each case the CSAT3 was used for the covariance/cospectra, with the anemometer's sample rate adjusted to match the sample rate of the hygrometer.

References

Alfieri, J.G., Kustas, W.P., Prueger, J.H., Hipps, L.E., Chávez, J.L., French, A.N., Evett, S. R., 2011. Intercomparison of Nine Micrometeorological Stations during the BEAREX08 Field Campaign. J. Atmos. Oceanic Technol. 28 (11), 1390–1406. https://doi.org/10.1175/2011jtech1514.1.

Amiro, B.D., 2001. Paired-tower measurements of carbon and energy fluxes following disturbance in the boreal forest [10.1046/j.1365-2486.2001.00398.x] Glob. Chang. Biol. 7 (3), 253–268. https://doi.org/10.1046/j.1365-2486.2001.00398.x.

Anthoni, P.M., Unsworth, M.H., Law, B.E., Irvine, J., Baldocchi, D.D., Tuyl, S.V., Moore, D., 2002. Seasonal differences in carbon and water vapor exchange in young and old-growth ponderosa pine ecosystems. Agric. For. Meteorol. 111 (3), 203–222. https://doi.org/10.1016/S0168-1923(02)00021-7.

Barr, J.G., Engel, V., Smith, T.J., Fuentes, J.D., 2012. Hurricane disturbance and recovery of energy balance, CO2 fluxes and canopy structure in a mangrove forest of the Florida Everglades. Agric. For. Meteorol. 153, 54–66. https://doi.org/10.1016/j.agrformet.2011.07.022.

Beaton, S.P., Spowart, M., 2012. UV absorption hygrometer for fast-response airborne water vapor measurements. J. Atmos. Oceanic Technol. 29 (9), 1295–1303.

Burba, G.G., McDermitt, D.K., Grelle, A., Anderson, D.J., Xu, L.K., 2008. Addressing the influence of instrument surface heat exchange on the measurements of CO2 flux from open-path gas analyzers [Article] Glob. Chang. Biol. 14 (8), 1854–1876. https://doi.org/10.1111/j.1365-2486.2008.01606.x.

Burba, G.G., McDermitt, D.K., Anderson, D.J., Furtaw, M.D., Eckles, R.D., 2010. Novel design of an enclosed CO2/H2O gas analyser for eddy covariance flux measurements. Chem. Phys. Meteorol. 62 (5), 743–748. https://doi.org/10.1111/j.1600-0889.2010.00468.x.

Burns, S.P., Frank, J.M., Massman, W.J., Patton, E.G., Blanken, P.D., 2021. The effect of static pressure-wind covariance on vertical carbon dioxide exchange at a windy subalpine forest site. Agric. For. Meteorol. 306, 108402 https://doi.org/10.1016/j.agrformet.2021.108402.

Charuchittipan, D., Babel, W., Mauder, M., Leps, J.-.P., Foken, T., 2014. Extension of the Averaging Time in Eddy-Covariance Measurements and Its Effect on the Energy Balance Closure. Boundary Layer Meteorol. 152 (3), 303–327. https://doi.org/10.1007/s10546-014-9922-6.

Eshonkulov, R., Poyda, A., Ingwersen, J., Pulatov, A., Streck, T., 2019. Improving the energy balance closure over a winter wheat field by accounting for minor storage terms. Agric. For. Meteorol. 264, 283–296. https://doi.org/10.1016/j.agrformet.2018.10.012.

Frank, J.M., Massman, W.J., Ewers, B.E., 2013. Underestimates of sensible heat flux due to vertical velocity measurement errors in non-orthogonal sonic anemometers. Agric. For. Meteorol. 171–172, 72–81. https://doi.org/10.1016/j.agrformet.2012.11.005 (0).

Frank, J.M., Massman, W.J., Ewers, B.E., Huckaby, L.S., Negrón, J.F., 2014. Ecosystem CO_2/H_2O fluxes are explained by hydraulically limited gas exchange during tree mortality from spruce bark beetles. J. Geophys. Res. 119 (6), 1195–1215.

Frank, J.M., Massman, W.J., Swiatek, E., Zimmerman, H.A., Ewers, B.E., 2016a. All sonic anemometers need to correct for transducer and structural shadowing in their velocity measurements. J. Atmos. Oceanic Technol. 33, 149–167. https://doi.org/10.1175/jtech-D-15-0171.1.

Frank, J.M., Massman, W.J., Ewers, B.E., 2016b. A Bayesian model to correct underestimated 3-D wind speeds from sonic anemometers increases turbulent components of the surface energy balance. Atmos. Meas. Tech. 9 (12), 5933–5953. https://doi.org/10.5194/amt-9-5933-2016.

Frank, J.M., Massman, W.J., Ewers, B.E., Williams, D.G., 2019. Bayesian analyses of 17 winters of water vapor fluxes show bark beetles reduce sublimation. Water Resour. Res. 55 (2), 1598–1623.

Frank, J.M., Massman, W.J., 2020. A new perspective on the open-path infrared gas analyzer self-heating correction. Agric. For. Meteorol. 290, 107986 https://doi.org/10.1016/j.agrformet.2020.107986.

Frank, J.M., Massman, W.J, 2021. AmeriFlux BASE US-GLE GLEES, Ver. 8-5. AmeriFlux AMP. https://doi.org/10.17190/AMF/1246056 (Dataset).

Fratini, G., Ibrom, A., Arriga, N., Burba, G., Papale, D., 2012. Relative humidity effects on water vapour fluxes measured with closed-path eddy-covariance systems with short sampling lines. Agric. For. Meteorol. 165, 53–63. https://doi.org/10.1016/j.agrformet.2012.05.018.

Fratini, G., McDermitt, D.K., Papale, D., 2014. Eddy-covariance flux errors due to biases in gas concentration measurements: origins, quantification and correction. Biogeosciences 11 (4), 1037–1051. https://doi.org/10.5194/bg-11-1037-2014.

Ha, W., Kolb, T.E., Springer, A.E., Dore, S., O'Donnell, F.C., Martinez Morales, R., Koch, G.W., 2015. Evapotranspiration comparisons between eddy covariance measurements and meteorological and remote-sensing-based models in disturbed ponderosa pine forests [10.1002/eco.1586] Ecohydrology 8 (7), 1335–1350. https://doi.org/10.1002/eco.1586.

Haslwanter, A., Hammerle, A., Wohlfahrt, G., 2009. Open-path vs. closed-path eddy covariance measurements of the net ecosystem carbon dioxide and water vapour exchange: a long-term perspective. Agric. For. Meteorol. 149 (2), 291–302.
Haverd, V., Cuntz, M., Leuning, R., Keith, H., 2007. Air and biomass heat storage fluxes in a forest canopy: calculation within a soil vegetation atmosphere transfer model [Article] Agric. For. Meteorol. 147 (3–4), 125–139. https://doi.org/10.1016/j.agrformet.2007.07.006.
Horst, T., Semmer, S., Maclean, G., 2015. Correction of a non-orthogonal, three-component sonic anemometer for flow distortion by transducer shadowing. Boundary Layer Meteorol. 155 (3), 371–395.
Ibrom, A., Dellwik, E., Flyvbjerg, H., Jensen, N.O., Pilegaard, K., 2007. Strong low-pass filtering effects on water vapour flux measurements with closed-path eddy correlation systems. Agric. For. Meteorol. 147 (3), 140–156. https://doi.org/10.1016/j.agrformet.2007.07.007.
Irmak, S., Payero, J.O., Kilic, A., Odhiambo, L.O., Rudnick, D., Sharma, V., Billesbach, D., 2014. On the magnitude and dynamics of eddy covariance system residual energy (energy balance closure error) in subsurface drip-irrigated maize field during growing and non-growing (dormant) seasons. Irrigation Sci. 32 (6), 471–483. https://doi.org/10.1007/s00271-014-0443-3.
Judd, M.J., Prendergast, P.T., McAneney, K.J., 1993. Carbon dioxide and latent heat flux measurements in a windbreak-sheltered orchard. Agric. For. Meteorol. 66 (3), 193–210. https://doi.org/10.1016/0168-1923(93)90071-O.
Kochendorfer, J., Meyers, T.P., Frank, J.M., Massman, W.J., Heuer, M.W., 2012. How well can we measure the vertical wind speed? Implications for fluxes of energy and mass. Boundary Layer Meteorol. 145 (2), 383–398. https://doi.org/10.1007/s10546-012-9738-1.
Kosugi, Y., Takanashi, S., Tanaka, H., Ohkubo, S., Tani, M., Yano, M., Katayama, T., 2007. Evapotranspiration over a Japanese cypress forest. I. Eddy covariance fluxes and surface conductance characteristics for 3years. J. Hydrol. (Amst) 337 (3), 269–283. https://doi.org/10.1016/j.jhydrol.2007.01.039.
Kristensen, L., Mann, J., Oncley, S.P., Wyngaard, J.C., 1997. How close is close enough when measuring scalar fluxes with displaced sensors? J. Atmos. Oceanic Technol. 14 (4), 814–821. https://doi.org/10.1175/1520-0426(1997)014<0814:hcicew>2.0.co;2.
Kruschke, J, 2014. Doing Bayesian data analysis: A tutorial With R. Academic Press, JAGS, and Stan.
Kutikoff, S., Lin, X., Evett, S.R., Gowda, P., Brauer, D., Moorhead, J., Owensby, C., 2021. Water vapor density and turbulent fluxes from three generations of infrared gas analyzers. Atmos. Meas. Tech. 14 (2), 1253–1266. https://doi.org/10.5194/amt-14-1253-2021.
Lee, X., Black, T.A., Novak, M.D., 1994. Comparison of flux measurements with open- and closed-path gas analyzers above an agricultural field and a forest floor. Boundary Layer Meteorol. 67 (1), 195–202. https://doi.org/10.1007/bf00705514.
Leuning, R., King, K., 1992. Comparison of eddy-covariance measurements of CO_2 fluxes by open-and closed-path CO_2 analysers. Boundary Layer Meteorol. 59 (3), 297–311.
Leuning, R., van Gorsel, E., Massman, W.J., Isaac, P.R., 2012. Reflections on the surface energy imbalance problem. Agric. For. Meteorol. 156, 65–74. https://doi.org/10.1016/j.agrformet.2011.12.002.
Loescher, H.W., Hanson, C.V., Ocheltree, T.W., 2009. The psychrometric constant is not constant: a novel approach to enhance the accuracy and precision of latent energy fluxes through automated water vapor calibrations [Article] J. Hydrometeorol. 10 (5), 1271–1284. https://doi.org/10.1175/2009jhm1148.1.
Lu, D., Ye, M., Hill, M.C., 2012. Analysis of regression confidence intervals and Bayesian credible intervals for uncertainty quantification. Water Resour. Res. 48 (9), W09521. https://doi.org/10.1029/2011wr011289.
Mammarella, I., Launiainen, S., Gronholm, T., Keronen, P., Pumpanen, J., Rannik, Ü., Vesala, T., 2009. Relative Humidity Effect on the High-Frequency Attenuation of Water Vapor Flux Measured by a Closed-Path Eddy Covariance System. J. Atmos. Oceanic Technol. 26 (9), 1856–1866. https://doi.org/10.1175/2009jtecha1179.1.
Martínez-Cob, A., Suvočarev, K., 2015. Uncertainty due to hygrometer sensor in eddy covariance latent heat flux measurements. Agric. For. Meteorol. 200, 92–96. https://doi.org/10.1016/j.agrformet.2014.09.021.
Massman, W.J., 1991. The attenuation of concentration fluctuations in turbulent-flow through a tube. J. Geophys. Res. 96 (D8), 15269–15273. https://doi.org/10.1029/91jd01514.
Massman, W.J., 1999. Molecular diffusivities of Hg vapor in air, O_2 and N_2 near STP and the kinematic viscosity and thermal diffusivity of air near STP. Atmos. Environ. 33 (3), 453–457. https://doi.org/10.1016/s1352-2310(98)00204-0.
Massman, W.J., 2000. A simple method for estimating frequency response corrections for eddy covariance systems [Article] Agric. For. Meteorol. 104 (3), 185–198. https://doi.org/10.1016/s0168-1923(00)00164-7.
Massman, W.J., Ibrom, A., 2008. Attenuation of concentration fluctuations of water vapor and other trace gases in turbulent tube flow. Atmos. Chem. Phys. 8 (20), 6245–6259.
MATLAB, 2018. Version 9.5.0.944444 (R2018b). The MathWorks, Inc, Natick, Massachusetts.
Mauder, M., Liebethal, C., Göckede, M., Leps, J.-P., Beyrich, F., Foken, T., 2006. Processing and quality control of flux data during LITFASS-2003. Boundary Layer Meteorol. 121 (1), 67–88. https://doi.org/10.1007/s10546-006-9094-0.
Mauder, M., Oncley, S.P., Vogt, R., Weidinger, T., Ribeiro, L., Bernhofer, C., Liu, H., 2007. The energy balance experiment EBEX-2000. Part II: intercomparison of eddy-covariance sensors and post-field data processing methods [Article] Boundary Layer Meteorol. 123 (1), 29–54. https://doi.org/10.1007/s10546-006-9139-4.
Mauder, M., Foken, T., Cuxart, J., 2020. Surface-Energy-Balance Closure over Land: a Review. Boundary Layer Meteorol. 177 (2), 395–426. https://doi.org/10.1007/s10546-020-00529-6.
Moorhead, J.E., Marek, G.W., Gowda, P.H., Lin, X., Colaizzi, P.D., Evett, S.R., Kutikoff, S., 2019. Evaluation of Evapotranspiration from Eddy Covariance Using Large Weighing Lysimeters. Agronomy 9 (2), 99.
Morrison, T., Pardyjak, E.R., Mauder, M., Calaf, M., 2022. The Heat-Flux Imbalance: the Role of Advection and Dispersive Fluxes on Heat Transport Over Thermally Heterogeneous Terrain. Boundary Layer Meteorol. 183 (2), 227–247. https://doi.org/10.1007/s10546-021-00687-1.
Musselman, R.C., 1994. The Glacier Lakes Ecosystem Experiments Site. (General Technical Report RM–249). Fort Collins, CO: USDA Forest Service, Rocky Mountain Forest and Range Experiment Station.
Novick, K., Walker, J., Chan, W., Schmidt, A., Sobek, C., Vose, J., 2013. Eddy covariance measurements with a new fast-response, enclosed-path analyzer: spectral characteristics and cross-system comparisons. Agric. For. Meteorol. 181, 17–32.
Odum, H.T., 1957. Trophic structure and productivity of Silver Springs. Florida. Ecological Monographs 27 (1), 55–112. https://doi.org/10.2307/1948571.
Plummer, M., 2003. JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. In: Paper presented at the Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003) March 20–22. Vienna, Austria.
Plummer, M., 2022. Rjags: Bayesian Graphical Models using MCMC. R package version 4-13. https://CRAN.R-project.org/package=rjags.
Plummer, M., Best, N., Cowles, K., Vines, K., 2006. CODA: convergence diagnosis and output analysis for MCMC. R. News 6, 7–11.
Polonik, P., Chan, W., Billesbach, D., Burba, G., Li, J., Nottrott, A., Biraud, S., 2019. Comparison of gas analyzers for eddy covariance: effects of analyzer type and spectral corrections on fluxes. Agric. For. Meteorol. 272, 128–142.
R. Core Team, 2015. R: a language and environment for statistical computing.
Reed, D.E., Frank, J.M., Ewers, B.E., Desai, A.R., 2018. Time dependency of eddy covariance site energy balance. Agric. For. Meteorol. 249, 467–478. https://doi.org/10.1016/j.agrformet.2017.08.008.
Rstudio Team, 2015. Rstudio: integrated Development for R.
Runkle, B.R.K., Wille, C., Gažovič, M., Kutzbach, L., 2012. Attenuation correction procedures for water vapour fluxes from closed-path eddy-covariance systems. Boundary Layer Meteorol. 142 (3), 401–423. https://doi.org/10.1007/s10546-011-9689-y.
Ryken, A.C., Gochis, D., Maxwell, R.M., 2022. Unravelling groundwater contributions to evapotranspiration and constraining water fluxes in a high-elevation catchment. Hydrol. Process 36 (1), e14449. https://doi.org/10.1002/hyp.14449.
Smith, R.J., 2009. Use and misuse of the reduced major axis for line-fitting. Am. J. Phys. Anthropol. 140 (3), 476–486. https://doi.org/10.1002/ajpa.21090.
Sonntag, D., Foken, T., Vömel, H., & Hellmuth, O., 2021. Humidity sensors. Springer handbook of atmospheric measurements, 209–241.
Stoy, P.C., Mauder, M., Foken, T., Marcolla, B., Boegh, E., Ibrom, A., Varlagin, A., 2013. A data-driven analysis of energy balance closure across FLUXNET research sites: the role of landscape scale heterogeneity. Agric. For. Meteorol. 137–152. https://doi.org/10.1016/j.agrformet.2012.11.004, 171–172(0).
Twine, T.E., Kustas, W.P., Norman, J.M., Cook, D.R., Houser, P.R., Meyers, T.P., Wesely, M.L., 2000. Correcting eddy-covariance flux underestimates over a grassland [Article] Agric. For. Meteorol. 103 (3), 279–300. https://doi.org/10.1016/s0168-1923(00)00123-4.
Webb, E.K., Pearman, G.I., Leuning, R., 1980. Correction of flux measurements for density effects due to heat and water vapour transfer [Article] Q. J. R. Meteorolog. Soc. 106 (447), 85–100. https://doi.org/10.1256/smsqj.44706.
Wilson, K., Goldstein, A., Falge, E., Aubinet, M., Baldocchi, D., Berbigier, P., Verma, S., 2002. Energy balance closure at FLUXNET sites. Agric. For. Meteorol. 113 (1–4), 223–243. https://doi.org/10.1016/s0168-1923(02)00109-0.
Wu, J.B., Zhou, X.Y., Wang, A.Z., Yuan, F.H., 2015. Comparative measurements of water vapor fluxes over a tall forest using open- and closed-path eddy covariance system. Atmos. Meas. Tech. 8 (10), 4123–4131. https://doi.org/10.5194/amt-8-4123-2015.
Wyngaard, J.C., Zhang, S.-F., 1985. Transducer-shadow effects on turbulence spectra measured by sonic anemometers. J. Atmos. Oceanic Technol. 2 (4), 548–558. https://doi.org/10.1175/1520-0426(1985)002<0548:tseots>2.0.co;2.
van Dijk, A., Kohsiek, W., de Bruin, H.A.R., 2003. Oxygen sensitivity of krypton and Lyman-α hygrometers. J. Atmos. Oceanic Technol. 20 (1), 143–151. https://doi.org/10.1175/1520-0426(2003)020<0143:osokal>2.0.co;2.
Zhang, W., Nelson, J.A., Poyatos, R., Miralles, D., Migliavacca, M., Reichstein, M., & Jung, M., 2021. Improved eddy covariance flux based transpiration estimates at high relative humidity and comparison to sap flux. Paper presented at the EGU General Assembly Conference Abstracts.