Determining Meaningful Differences for SMACEX Eddy Covariance Measurements

D. W. MEek AND J. H. PRueGHER
National Soil Tilth Laboratory, ARS, USDA, Ames, Iowa

W. P. KUSTAS
Hydrology and Remote Sensing Laboratory, ARS, USDA, Beltsville, Maryland

J. L. HAIfIELD
National Soil Tilth Laboratory, ARS, USDA, Ames, Iowa

(Manuscript received 1 May 2004, in final form 4 March 2005)

ABSTRACT

Two eddy covariance instrument comparison studies were conducted before and after the Soil Moisture—Atmosphere Coupling Experiment (SMACEX) field campaign to 1) determine if observations from multiple sensors were equivalent for the measured variables over a uniform surface and to 2) determine a least significant difference (LSD) value for each variable to discriminate between daily and hourly differences in latent and sensible heat and carbon dioxide fluxes, friction velocity, and standard deviation of the vertical wind velocity from eddy covariance instruments placed in different locations within the study area. The studies were conducted in early June over an alfalfa field and in mid-September over a short grass field. Several statistical exploratory, graphical, and multiple-comparison procedures were used to evaluate each daily variable. Daily total or average data were used to estimate a pooled standard error and corresponding LSD values at the $P = 0.05$ and $P = 0.01$ levels using univariate procedures. There were no significant sensor differences in any of the daily measurements for either intercomparison period. Hourly averaged data were used to estimate a pooled standard error and corresponding LSD values at the $P = 0.05$ and $P = 0.01$ levels using mixed model procedures. Sensor differences for pre- and post-intercomparison were minimal for hourly and daily values of CO$_2$, water vapor, sensible heat, friction velocity, and standard deviation for vertical wind velocity. Computed LSD values were used to determine significant daily differences and threshold values for the variables monitored during the SMACEX campaign.

1. Introduction

Eddy covariance (EC) is a micrometeorological technique used to measure turbulence exchange flux in the vertical of mass, momentum, and energy between a surface and the boundary layer of the atmosphere. Increased availability and affordability of sonic anemometers and infrared gas analyzers (IRGA) have resulted in a wide application of this technique in many surface energy balance studies. Multiple EC systems (Spittlehouse and Black 1979, 1980; Shuttleworth et al. 1988), deployed over a range of heterogeneous or homogeneous landscapes, raise an important question as to determining when a flux (mass, energy, or momentum) from one surface is significantly different than that from another surface. The intercomparison of multiple eddy covariance systems over a common surface provides a means of assessing instrument bias and establishing threshold levels of meaningful differences for fluxes when evaluating EC systems over spatially distributed surfaces.

Fritschen et al. (1992) compared sensible and latent heat fluxes ($H$ and $L.E.$ respectively) from three different types of Bowen ratio (BR) and two EC systems used in the First International Satellite Land Surface Climatology Project (ISLSCP) Field Experiment (FIFE) in 1989, and reported close agreement among the BR types but less satisfactory comparisons between the EC and BR systems. Intercomparison between the two EC systems was not reported. E. T. Kanemasu et al. (1987, personal communication) compared the BR and EC approaches and found large discrepancies between techniques and instruments. Similar results were also reported by Spittlehouse and Black (1979, 1980)
and Shuttleworth et al. (1988). Nie et al. (1992) reported that, on a daily average basis, different instruments could account for up to a 20% difference for daily LE measurements and up to 90 W m\(^{-2}\) in LE for a single 30-min average. Similar to Fritschcn et al. (1992) these were differences between BR and EC systems, but not among individual EC systems. Tsvang et al. (1985) reported correlation coefficients in the range of 0.85–0.90 for turbulence characteristics from various types of sonic anemometers, thermoanemometers, and resistance thermometers. The consensus is that variation within an instrument limits the achievable accuracy of flux estimates.

Carbon dioxide (CO\(_2\)) flux measurements using eddy covariance have become an important component of many surface energy balance studies and, thus, naturally raise the question of instrument variability. An intercomparison study that included different CO\(_2\) sensors used with eddy covariance during FIFE 1989 (Moncrief et al. 1992) showed CO\(_2\) flux agreement to within 15%. They also reported that models of the same type of CO\(_2\) sensor agreed to within 5%. A common recommendation mentioned in nearly all past intercomparison studies was the need to conduct intercomparisons as a means to assess and evaluate the instrument system bias. This is particularly important when using multiple eddy covariance systems to address spatial and temporal surface energy balance questions.

The Soil Moisture–Atmosphere Exchange Experiment (SMACEX) was an interdisciplinary investigation involving a diverse set of field measurements and modeling activities funded by the National Aeronautics and Space Administration (NASA) Terrestrial Hydrology Program and the U.S. Department of Agriculture (USDA) Agricultural Research Service (ARS). Measurements of the coupled exchange of water vapor, carbon dioxide, and energy fluxes between the soil, vegetation, and atmosphere included 12 tower-based EC systems, with 6 in corn- (Zea mays L.) and 6 in soybean (Glycine max L.) fields. A critical component of the SMACEX field campaign included two EC intercomparison studies (Kustas et al. 2005). The purpose of these intercomparisons was twofold. The first goal was to determine if all the EC sensors produce reasonably equivalent daily and hourly estimates for each variable of interest in the SMACEX study. The second goal was to determine a common significant difference for daily and hourly measurements for fluxes of water vapor (LE), sensible heat (H), CO\(_2\), friction velocity (\(u_\ast\)), and the standard deviation of vertical wind velocity (\(s_{u_v}\)) when any two sensors were placed in similar or contrasting environments (i.e., 6 cornfields and 6 soybean fields) during the SMACEX observation period.

### 2. Collocation sites and instrumentation

The first intercomparison was conducted prior to the SMACEX study; the second was immediately after the end of the SMACELX study. These are denoted as the pre-SMACEX and post-SMACEX periods, respectively.

#### a. Instrumentation

Twelve EC systems were compared in the intercomparison study that was conducted in central Iowa. Each EC system was composed of a three-dimensional sonic anemometer (CSAT3, Campbell Scientific, Inc. (CSI), Logan, Utah), with 10 CSAT3s paired with a fast-response water vapor (H\(_2\)O) and CO\(_2\) density open-path IRGA (L17500, LI-COR, Inc., Lincoln, Nebraska) and the remaining two CSAT3s paired with a water vapor–only sensor (KH20 krypton hygrometer, CSI). These units were positioned 1.7 m above the surface and were oriented in a south-facing direction with approximately 2-m separation among individual pairs of EC units along an east–west transect. Upwind fetch of the prevailing wind direction (south) was in excess of 200 m. Each individual EC system was connected to a separate Campbell Scientific datalogger (CR23x) to record and store the data. The sampling frequency for all systems was 20 Hz.

#### b. Pre-SMACEEX intercomparison

The location of the pre-SMACEX intercomparison was in a 21-ha field [41.957°N, 93.609°W, 350 m above sea level (ASL)] with no relief (i.e., flat surface). Observations of all instruments began at 1240 local standard time (LST) on day of year (DOY) 156 and continued through 0630 LST on DOY 157 of 2002 over an alfalfa (Medicago sativa L.) canopy. The alfalfa had been recently harvested, and regrowth had not yet completely covered the ground surface. This was the most uniform vegetative surface available in the area at this time, and it provided a surface with similar ground cover and adequate latent heat (LE) fluxes to compare differences among the instruments. Meteorological conditions were typical of this period, with partly cloudy skies, temperatures that ranged from 10° to 22°C, and wind speeds averaging 1.8 m s\(^{-1}\). This set had data for 0.75 days recorded every 20 min in which all of the instruments were present.

---

1The mention of trade names or products is solely for the purpose of providing specific information and does not imply endorsement by the USDA.
c. Post-SMACEX intercomparison

The location of this comparison was in a 4-ha field (42.023°N, 93.774°W, 353.6 m ASL), also with little relief. Data collection here began at 1100 LST on DOY 259 and continued through 0830 LST on DOY 261. The instruments were placed over a surface of short (0.10 m)-fescue smooth-brome grass sward (Bromus incernis Leyss). The grass was newly cut on the morning of the first day of the intercomparison. Meteorological conditions during the post-intercomparison were sunny, with temperatures that ranged from 10° to 27°C and wind speeds averaging 2 m s⁻¹, with wind directions out of the south and southwest. This set had 1.92 days recorded every 30 min in which all instruments were present. Note that both sets started at midday and recorded continuously to the end of each intercomparison period.

3. Statistical methods

Inspection of the data, along with the use of exploratory data analyses (EDA; Hoagland et al. 1983) and diagnostics, revealed some major concerns. These include the following: because of the before-and-after nature of the intercomparison, along with operational constraints and difficulties, the data record was serially incomplete for some variables (with some having the entire record missing on given days); all variables were autocorrelated within each collection period (see Jones 1975 and Meek et al. 1999 for related analyses issues); severe outliers were present; and for all of the variables the variability between the periods was unequal. The factors affected both the selection of data subsets and the methods of analyses used.

a. Graphical comparisons of time series

Each variable was examined for the pre- and post-SMACEX periods using either the hourly or daily composites. Several EDA multiple comparison procedures were used, including correlation scatterplot matrices, notched box plots (McGill et al. 1978) for median comparisons, mean assessments with graphical multiple comparisons using comparison circles (Sall 1992), including similar analysis on ranked data (e.g., Conover and Iman 1981), and overlain time series with all sensor responses on a common set of axes.

b. Statistical comparison of time series

The second goal was to assess whether statistically significant differences among instrument values could be detected among fields with different or similar crop and management practices. To complete these analyses, LE and H values were converted to megajoule per square meter and were totaled over the day, while CO₂ was totaled to milligrams per square meter. Friction velocity (u*) and the standard deviation of the vertical velocity (σᵥ) were averaged. Three defined periods were used in these analyses—the first one used all of the data from the pre-SMACEX period, and the second was the first complete 24 h of observations in the post-SMACEX period, while the third was the remaining observations in this set. Standard errors were computed for each period and weighted pooled standard errors were then computed from period values. The weights were the ratio of the individual period length to the sum of all of the period lengths. Finally, for each variable, the pooled standard error was used to estimate a least significant difference (LSD) at the P = 0.05 and P = 0.01 levels.

Evaluation of the hourly values of the variables provides a comparison of the differences at temporal resolutions for less than daily values. These analyses were conducted on observations during the daytime when net radiation is positive, which was between 0800 and 1800 LST for central Iowa for this time period. Data for CO₂, H, LE, u*, and σᵥ were averaged for each hourly period. This allows all variables from both periods to be examined on the same time scale so that the standard error can be readily pooled. Hourly datasets were constructed from serially complete subsets of the raw observations. There were 3 days with sufficient data to conduct these comparisons. For the pre-SMACEX datasets, there were 6–7 h of observations on DOY 156; for the post-SMACEX datasets, there were 7–8 h of observations on DOY 259 and 9–11 h for DOY 260. With the 3 days of data for five variables, 15 initial datasets were prepared.

An analysis of covariance (ANCOVA) was used to examine each variable. For each given variable y(hr), the sensor hourly response over time can be modeled as

\[ y(hr) = m_{sensor} + f(hr) + e(hr), \]

where \( m_{sensor} \) is the mean sensor response, \( f(hr) \) is a time trend common to all sensors, and \( e(hr) \) is the error term. The error term may be first-order autoregressive, and \( f(hr) \) is a first- to fourth-degree polynomial [Eq. (1)]. The analysis procedure followed the general methods given in Littell et al. (1996); here, \( m_{sensor} \) was treated as a random effect, \( f(hr) \) as a covariate, and \( e(hr) \), if needed, as a first-order autoregressive error. Hourly variances for each variable were pooled and weighted to determine the overall uncertainty estimate and LSD values.
4. Results

a. Graphical comparisons of time series

Many graphical, EDA, and other procedures were conducted, but for brevity only the most critical findings are reported. A montage of time series plots for all variables on DOY 259 of the post-SMACEX intercomparison is shown in Fig. 1 as an example of general diurnal response trends and sensor variability. Screening limits were the mean of sensor observations at each
time plus or minus three pooled standard deviations, where the standard deviations were computed and pooled separately over daytime and nighttime hours for each variable. In all cases the daytime variability was larger than the nighttime variability.

Insight was provided about the variation among sensors by evaluating sensor statistics within daily or hourly intervals. Variation of the hourly sensor values was compared among all daily periods for each variable. For illustration, we selected two sets of original observations from the daily results that exhibited the highest and lowest daily variability. The results are shown from EDA for the highest (CO₂, Fig. 2a) and lowest (σₚ, Fig. 2b) variability with box-whisker diagrams. The plot summarizes each sensor's central tendency and distribution, where the box width is the usual middle half (50%) of the data, while the whiskers cover the middle 90%. Within the box the median value (50th percentile) is represented as a thick line. On the right side of each plot the comparison circles represent the Tukey-Kramer intercomparison, as described by Sall (1992). Overlapping circles suggest equivalence while separated circles suggest distinction among values. For these variables, with the highest and least variability, no individual sensor was completely distinct from any other one or the group as a whole (Fig. 2). The visual comparisons from the graphical tests revealed that each of the variable comparisons were reasonably equivalent among all sensors for each daily period.

b. Statistical comparison of time series

The variability in the H and σₚ observations was larger for the pre- than for the post-SMACEX period. The opposite condition was true for the remaining variables. The coefficient of variation (CV) values ranged from 13.2% to 26.9% for LE, from 2.6% to 15.2% for H, from 15.8% to 61.6% for CO₂, from 5.3% to 8.3% for u*, and from 2.2% to 6.9% for σₚ. Most of the daily values were normally distributed based on two or more standard normality tests, but a few sets had at least one sensor falling below or above the end of a box-plot whisker (5th or 95th percentile). Table 1 reports the estimated daily total instrument standard error and resulting minimally acceptable instrument differences based on LSDₚₐ and LSD₀₀ for LE, H, and CO₂. These estimates mainly represent each sensor's daytime response during positive Rₑ periods when the sensor variability is larger and Rₑ values usually exceeded 50 W m⁻². The differences that we observed across the sites within SMACEX were often much larger than these minimum values (Prueger et al. 2005). The values for significant differences of u* and σₚ were also quite small in the intercomparison and were easily exceeded many times during the field campaign. These intercomparisons of instruments provide confidence that for each variable any difference that we observed between any 2 of the 12 cropped fields was a true difference, induced by an unlikeness in crop, soil, soil water status, and so forth.

The ANCOVA procedure that was used allows for trends and autocorrelation. Moreover, it allows the sensors to be biased with respect to one another, yet provides a variability estimate. To effectively use this procedure, some data screening and selection was required.

![Fig. 2. Selected exploratory comparisons on two sets of sensor observations over a day: (a) y represents CO₂ observations by a sensor for DOY 261, and (b) y represents σₚ by a sensor for DOY 261.](image)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Units</th>
<th>Pooled SE</th>
<th>LSD₀₀</th>
<th>LSDₚₐ</th>
</tr>
</thead>
<tbody>
<tr>
<td>LE</td>
<td>MJ m⁻²</td>
<td>0.59</td>
<td>1.28</td>
<td>1.80</td>
</tr>
<tr>
<td>H</td>
<td>MJ m⁻²</td>
<td>0.19</td>
<td>0.41</td>
<td>0.57</td>
</tr>
<tr>
<td>CO₂</td>
<td>g m⁻²</td>
<td>1.28</td>
<td>2.90</td>
<td>4.16</td>
</tr>
<tr>
<td>u*</td>
<td>m s⁻¹</td>
<td>0.012</td>
<td>0.027</td>
<td>0.038</td>
</tr>
<tr>
<td>σₚ</td>
<td>m s⁻¹</td>
<td>0.014</td>
<td>0.031</td>
<td>0.043</td>
</tr>
</tbody>
</table>
Instruments and sites within a region should be appropriately considered. Some questions that need to be answered include the representation of an instrument observation and the variation in the instrument footprint with respect to time, space, and ambient conditions. A network of instruments within a field may improve our understanding of any parameter's variability, however, the certainty of the values will be critical to being able to effectively interpret these results.

For both the daily and hourly values, there were no highly significant ($P < 0.01$) sensor differences in any of the measurements for either the pre- or post-SMACEX period. LSD values were estimated and used to determine the statistical differences among instruments in LE, $H$, $u^*$, $CO_2$, or $s_v$ values from the period of observation over similar surfaces; these LSD values were then used to assess mass and energy differences among sites during the SMACEX campaign (Prueger et al. 2005).

Acknowledgments. The authors are deeply appreciative of the efforts of Tim Hart and Forrest Goodman in the conduct of the field experiments. Their efforts to locate the sites and maintain the instruments during these intercomparisons were invaluable.

REFERENCES


Table 2. Discrimination values for hourly energy balance and turbulence components for pre- and post-intercomparisons.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Units</th>
<th>Pooled SE</th>
<th>LSD$_{0.05}$</th>
<th>LSD$_{0.01}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LE</td>
<td>W m$^{-2}$</td>
<td>9.8</td>
<td>24.6</td>
<td>38.0</td>
</tr>
<tr>
<td>$H$</td>
<td>W m$^{-2}$</td>
<td>13.4</td>
<td>29.2</td>
<td>41.1</td>
</tr>
<tr>
<td>$CO_2$</td>
<td>mg m$^{-2}$ s$^{-1}$</td>
<td>0.05</td>
<td>0.12</td>
<td>0.18</td>
</tr>
<tr>
<td>$u^*$</td>
<td>m s$^{-1}$</td>
<td>0.04</td>
<td>0.09</td>
<td>0.14</td>
</tr>
<tr>
<td>$s_v$</td>
<td>m s$^{-1}$</td>
<td>0.05</td>
<td>0.12</td>
<td>0.17</td>
</tr>
</tbody>
</table>

5. Discussion and conclusions

Our sensor comparison results for LE, $H$, and the two wind variables were similar to those reported in MacPherson et al. (1992), and were within the range of variability given by Goulden et al. (1996). For $CO_2$, our results were similar to Moncrieff et al.’s (1992), and also were within Goulden et al.’s. (1996) range rule. While regression slopes and CVs provide certain insights, there is no general standard method; thus, we recommend reporting pooled standard errors of similar instruments under similar conditions.

The LSD provides a robust method for point comparison between two instruments. Developing a suitable control chart is a graphical way to make and record such a comparison (e.g., Meek and Hatfield 2001). Additionally, the paired difference over time can be modeled (e.g., Meek et al. 2001). If multiple comparison hypotheses among several sites are of interest, appropriate procedures should be used (e.g., Westfall et al. 1999).

In future research efforts the variability issues among instruments and sites within a region should be thoroughly considered. Some questions that need to be answered include the representation of an instrument observation and the variation in the instrument footprint with respect to time, space, and ambient conditions. A network of instruments within a field may improve our understanding of any parameter’s variability, however, the certainty of the values will be critical to being able to effectively interpret these results.


