Uncertainty in Eddy Covariance Measurements: An Overview Based on a Recent Book Edited by Marc Aubinet, Timo Vesala, and Dario Papale

Aline Jaimes, Jaime Nava, and Vladik Kreinovich
Cyber-ShARE Center, University of Texas at El Paso
El Paso, Texas 79968, USA, contact vladik@utep.edu

1. Gap Filling

- Eddy covariance *algorithms* require that we have data from all moments of time.
- *In practice:* 20 to 60% of data points are faulty.
- *Examples:* wind measurements too high, or temperatures $> 20C$ degrees away from average.
- *Ideal:* get missing values from nearby meteostations.
- *If not possible:*
  - use interpolations (linear or nonlinear, e.g., NN) from before and after the gap,
  - from observations on the day before and day after,
  - use known relations between variables.
- *Result:* for short gaps, reconstructed values have almost the same accuracy as others.

2. Random and Systematic Error Components

- Measurement are never absolutely accurate; the result $\widetilde{x}$ is, in general, different from the actual value x.
- Moreover, if we repeatedly measure the same quantity x, we get slightly different values $\widetilde{x}_1, \ldots, \widetilde{x}_n$.
- Often, the measurement errors $\Delta x_i \stackrel{\text{def}}{=} \widetilde{x}_i -$ are purely "random", with equal prob. of $\Delta x_i > 0$ and $\Delta x_i < 0$.
- In this case, in the long run, these errors compensate each other, and we can estimate x as $\overline{x}_n \stackrel{\text{def}}{=} \dfrac{\widetilde{x}_1 + \ldots + \widetilde{x}_n}{n}$.
- The larger n, the more accurate this estimate $\overline{x}_n$.
- Sometimes, instruments have bias; in this case, $E[x] \stackrel{\text{def}}{=} \lim \overline{x}_n \neq x$.
- The difference $\Delta_s x \stackrel{\text{def}}{=} E[x] - x$ is called *systematic error*, and $\Delta_r x \stackrel{\text{def}}{=} \Delta x - \Delta_s x$ *random error*.

3. Random Errors of Eddy Covariance Measurements: Empirical Analysis

- In some places, there are two towers nearby, at similar places with practically the same carbon flux F.
- In this case, the difference between estimates $\widetilde{F}_1 - \widetilde{F}_2$ is caused only by the random error.
- So, we can get information about the random error by observing these differences.
- First observation: differences are normally distributed.
- Explanation:
  - *In general:* the measurement error is caused by many independent factors.
  - *Known:* sum of a large number of small indep. random variables has an almost normal distribution.

4. Empirical Analysis of Random Errors (cont-d)

- *Reminder:* random error is normally distributed.
- *Fact:* a normal distribution is uniquely determined by its mean μ and standard deviation σ.
- By definition of a random error, its mean is 0, so it is sufficient to know σ.
- How σ depends on F?
- In each measurement: the observation result $\widetilde{F}$ is slightly different from F.
- We can expand the dependence of $\widetilde{F}$ on F in Taylor series and ignore quadratic and higher order terms:

$$\widetilde{F} \approx a_0 + a_1 \cdot F.$$

- Measurement error can be caused both by the a_0-term and by the a_1-term.

5. Random Errors: Theoretical Analysis

- Reminder: Measurement error can be caused both by the a_0-term and by the a_1-term in the dependence

$$\widetilde{F} \approx a_0 + a_1 \cdot F.$$

  - the a_0-component does not depend on F while
  - the a_1-component is proportional to F.

- So, the random error has two components:
  - a component with $\sigma = a$ for some $a > 0$, and
  - component with $\sigma = b \cdot |F|$, for some b.

- These components are caused by the imperfection of the same measuring instrument.

- So, they must be strongly correlated.

- Hence, the overall error is $\sigma \approx a + b \cdot |F|$.

6. Random Errors: Comparison with Empirical Data

- *Theoretically:* the overall error σ depends on the flux value F as
$$\sigma \approx a + b \cdot |F|.$$
- *In the first approximation:* this is exactly what we observe.
- *The actual dependence* is somewhat more complex.
- *Specifically:*
  - when the flux F is small,
  - the observed error σ is smaller than the theoretical prediction $a + b \cdot |F|$.
- *It is not clear* what causes this difference.
- Explaining this empirical dependence is an interesting *challenge.*

7. Sources of Random Error

- *Reminder:* the computations of the flux are based on the theoretical model.
- *Assumption:* the parameters of the process do not change during the 30 minutes.
- *Resulting theoretical formula:* uses integration over the values at all spatial locations.
- *In practice:* we only have sensors at some locations.
- *As a result:* instead of average over all the values, we only average a sample.
- *Similar case:* to estimate the average weight, we only use a sample. In both cases, we have a random error.
- *Another source:* the footprint changes with time.
- *Yet another source:* sensors are not perfect.

8. Random Errors: Remaining Challenges

- *Fact:* we do not observe the random error Δx_i directly, we only observe the difference $\Delta x_1 - \Delta x_2$.
- *Good news:* if the distribution was symmetric, we can still reconstruct it from observing the difference.
- *Fact:* in general, the distribution is not symmetric.
- *Corollary:* we cannot distinguish the distribution for Δx and for $-\Delta x$.
- *Question:* what information can we recover?
  - for 1-D distribution, there is a huge uncertainty;
  - for 2-D case, e.g., for joint distribution of errors in two fluxes, the reconstruction is a.a. unique modulo

$$\Delta x \to -\Delta x.$$

- *Another problem:* reconstruction accuracy: skewness of Δx only affect the 6-th moment of $\Delta x_1 - \Delta x_2$.

9. Systematic Error

- *Two main sources* of systematic error:
  - approximate character of the model used in data processing, and
  - systematic error of sensors.
- *Main source of model inaccuracy:* non-turbulent fluxes are ignored.
- *In reality:* non-turbulent fluxes are not negligible, especially at night, when turbulence is lower.
- *First idea:* estimate the non-turbulent flux component.
- *Problem:* the Eddy covariance tower is not well suited for such estimation.
- *Result:* the "corrected" flux is less accurate than the original one.
- *Better solution:* ignore periods when turbulence is low.

10. Systematic Component of Instrument Error

- *Idea:* we can eliminate systematic error by calibration.
- *Notation:* let t_0 be the time of the calibration.
- *Problem:* the sensor starts shifting again.
- *Solution:* a new calibration is needed after some time T.
- *During the 2nd calibration:* we find the shift s of the sensor during the period from t_0 to $t_0 + T$.
- *Natural hypothesis:* since deviations are small, it is reasonable to assume that $\Delta_s(t)$ is a linear function of t:

$$\Delta_s(t) = k \cdot (t - t_0).$$

- We know that $\Delta_s(t_0 + T) = s$, so we can determine k from $s = k \cdot T$, as $k = \dfrac{s}{T}$.

11. How to Use Calibration to Improve the Accuracy of the Flux Estimation

- *Reminder:* it is reasonable to assume that

$$\Delta_s(t) = k \cdot (t - t_0), \text{ where:}$$

  - the value t_0 is the time of the previous calibration, and
  - the value s can be determined after the next calibration.

- *Idea:* after the calibration, we can correct the previous measurement results:
  - we subtract the systematic error $\Delta_s(t) = k \cdot (t - t_0)$ from all previous measurement results, and
  - by using the corrected values of all measured quantities, we get a more estimate estimate of the flux.

12. Systematic Error: Remaining Challenge

- *Systematic error:* even after calibration, there is a remaining part of the systematic error.
- *Estimates:* the un-corrected systematic error is gauged by providing an upper bound Δ_s such that $|\Delta_s x| \le \Delta_s$.
- *Fact:* systematic error in measuring instruments leads to a systematic error in flux F.
- *Question:* how to estimate the resulting systematic error in flux.
- *What we did:* Aline has already started these estimations.
- *Plan:* this is one of the main things on which we will concentrate.