Journal of Uncertain Systems
Vol.7, No.3, pp.181-185, 2013
Online at: www.jus.org.uk

How to Gauge Random Error in Eddy Covariance Measurements

Jaime Nava[1,2,*], Aline Jaimes[1], Craig Tweedie[1]
[1] *Cyber-ShARE Center, University of Texas at El Paso, El Paso, TX 79968, USA*
[2] *Hewlett Packard, El Paso, Texas, USA*

Received 1 April 2012; Revised 21 March 2013

Abstract

Random error is usually estimated in two ways: either by measuring the same quantity several times by the same measuring instrument or by measuring the same quantity using several similar measuring instruments. In both cases, we then compute the differences of the measurement results to estimate the random error. In environmental sciences, important information comes from eddy covariance measurements that measure time-changing fluxes of CO_2, H_2O, heat, etc. For these measurements, it is not easy to apply usual methods of estimating random error: we do not have a standard flux that can be used for calibration, and it is difficult to bring additional instruments to the tower for measuring the same flux. Thus we need new easier methods to estimate random errors of these flux measurements. For this estimation, we propose to use the fact that while the actual flux smoothly changes with time, the measurement results are non-smooth or "wiggly". This non-smoothness is caused by the difference between the measurement results and the actual values of the flux, i.e., by the random component of the measurement error. (The systematic component is the same for all the measurements and therefore does not introduce non-smoothness.) It is therefore reasonable to use the observed non-smoothness to estimate the random component of the measurement error.

Keywords: random error, eddy covariance measurements, environmental sciences

1 Formulation of the Problem

Measurement problem. To better understand meteorological and environmental processes, scientists need to estimate vertical exchanges of heat and different substances (such as CO_2 and H_2O) between the Earth surface and the atmosphere [1]. The exchange amount per unit of time and per unit of area is known as *flux*.

Most of this exchange is carried by turbulent atmospheric processes, mostly by *eddies* – rotating air-flows of different sizes. The corresponding fluxes can be measured by a tower equipped with measuring instruments at different heights; such a tower is called an *eddy covariance tower*. A single tower measures flow through a single point within the area of interest. To estimate the flux, i.e. the flow over the area of interest, scientists commonly integrate the flow values.

Need to estimate measurement uncertainty. The ultimate objective of flux measurement is to make important environmental decisions based on the results of these measurements. For example, if we observe that in a certain geographic area, plants start processing fewer CO_2, then something needs to be done to enhance the vitality of these plants and their ability to generate oxygen.

Since measurements are never absolutely accurate, to make meaningful decisions we need to know whether the change in the measured flux values reflects the actual flux change or the actual flux didn't change but the observed values are different because of the random measurement errors. To make this distinction, we need to know how big the random measurement errors can be, i.e., we need to estimate the random errors.

What is random error: reminder. The *measurement error* Δx is defined as the difference $\Delta x = \widetilde{x} - x$ between the measurement result $\widetilde{x}$ and the actual value x of the desired quantity; see, e.g., [2].

*Corresponding author. Email: nava.jaime@gmail.com (J. Nava).

Measurement results are never absolutely accurate and never absolutely reliable. A natural way to increase accuracy and reliability is make several repeated measurements. If we repeatedly use the same measuring instrument to measure the same quantity x, then, in general, we get slightly *different* measurement results – since while the instrument is the same and the quantity is the same, factors affecting accuracy may change. In other words, in this case, the measurement result can be viewed as a *random* variable, in which different values appear with different probability. Thus, the measurement error $\Delta x = \widetilde{x} - x$ is also a random variable.

The expected (mean) value of the measurement error is known as *systematic error component* (or *systematic error*, for short) $\Delta x_{\rm s} \stackrel{\rm def}{=} E[\Delta x]$, and the remaining part of the measurement error is called a *random error component* (or simply *random error*) $\Delta x_{\rm r} \stackrel{\rm def}{=} \Delta x - \Delta x_{\rm s}$.

By definition, the random error has zero mean:

$$E[\Delta x_{\rm r}] = E[\Delta x - \Delta x_{\rm s}] = E[\Delta x] - E[\Delta x_{\rm s}] = \Delta x_{\rm s} - \Delta x_{\rm s} = 0.$$

To characterize random error, usually, its standard deviation σ is used. In general, the standard deviation is defined as $\sigma^2 = E\left[(\Delta x_{\rm r} - E[\Delta x_{\rm r}])^2\right]$. Since $E[\Delta x_{\rm r}] = 0$, this formula can be simplified:

$$\sigma^2 = E\left[(\Delta x_{\rm r})^2\right] = E\left[(\Delta x - E[\Delta x])^2\right].$$

How random error is usually estimated. The problem of estimating random error is ubiquitous in measurement practice; see, e.g., [2]. In some measuring situations, we have a "standard" measuring instrument, i.e., an instrument which is much more accurate than the one whose random error we are estimating. In such situations, we can simply compare the results $\widetilde{x}$ obtained by our measuring instrument with the results $\widetilde{x}_{\rm st}$ obtained by the standard measuring instrument, and determine all the needed characteristics of the random error.

Specifically, since the measurement error $\Delta x_{\rm st} = \widetilde{x}_{\rm st} - x$ of the standard measuring instrument is much smaller than the measurement error Δx of our measuring instrument, we can safely ignore the value $\Delta x_{\rm st}$ and thus, take the result $\widetilde{x}_{\rm st}$ of applying the standard measuring instrument as the actual value x. In this approximation, the measurement error $\Delta x_i = \widetilde{x}_i - x$ of each measurement $i = 1, \ldots, n$ can be approximated as $\Delta x_i \approx \widetilde{x}_i - \widetilde{x}_{\rm st}$. Based on these values, we can estimate the desired standard deviation σ^2 in the usual way: first, we estimate the mean

$$\Delta x_{\rm s} \approx \frac{1}{n} \cdot \sum_{i=1}^{n} \Delta x_i,$$

and then, estimate the standard deviation by using the formula

$$\sigma^2 \approx \frac{1}{n-1} \cdot \sum_{i=1}^{n} (\Delta x_i - \Delta x_{\rm s})^2.$$

When the standard measuring instrument is not easily available, another way to gauge the random error is to repeatedly measure the same quantity by two similar measuring instruments. By definition of the measurement error components, the results $\widetilde{x}^{(1)}$ and $\widetilde{x}^{(2)}$ can be described as

$$\widetilde{x}^{(1)} = x + \Delta x_{\rm s}^{(1)} + \Delta x_{\rm r}^{(1)}; \quad \widetilde{x}^{(2)} = x + \Delta x_{\rm s}^{(2)} + \Delta x_{\rm r}^{(2)}.$$

Subtracting the expressions corresponding to the two instruments, we conclude that

$$\widetilde{x}^{(2)} - \widetilde{x}^{(1)} = \left(\Delta x_{\rm s}^{(2)} - \Delta x_{\rm s}^{(1)}\right) + \left(\Delta x_{\rm r}^{(2)} - \Delta x_{\rm r}^{(1)}\right).$$

Here, the mean value of the difference $\widetilde{x}^{(2)} - \widetilde{x}^{(1)}$ is equal to $\Delta x_{\rm s}^{(2)} - \Delta x_{\rm s}^{(1)}$. Since measurement errors corresponding to different measuring instruments are usually independent, it makes sense to assume that the random variables $\Delta x_{\rm r}^{(2)}$ and $\Delta x_{\rm r}^{(1)}$ are independent. Under this assumption, we can conclude that the variance of the difference $\widetilde{x}^{(2)} - \widetilde{x}^{(1)}$ is equal to the sum of the variances corresponding to $\Delta x_{\rm r}^{(2)}$ and $\Delta x_{\rm r}^{(1)}$. The two measuring instruments are similar, so they have the same standard deviation σ and variance σ^2; therefore, the standard deviation of the difference $\widetilde{x}^{(2)} - \widetilde{x}^{(1)}$ is equal to $2\sigma^2$.

Thus, if we have n pairs of results $\widetilde{x}_i^{(1)}$ and $\widetilde{x}_i^{(2)}$ of measuring the same quantities x_i by two similar measuring instruments, then we can estimate the standard deviation σ of the random error of each measuring instrument as follows. First, we estimate the mean difference $m \stackrel{\text{def}}{=} \Delta x_{\rm s}^{(2)} - \Delta x_{\rm s}^{(1)}$ between the systematic errors as the average difference between the measurement results:

$$m \approx \frac{1}{n} \cdot \sum_{i=1}^{n} \left(\widetilde{x}_i^{(2)} - \widetilde{x}_i^{(1)} \right);$$

after that, we estimate the variance $V = 2\sigma^2$ of the difference between the measurement results as

$$V = 2\sigma^2 \approx \frac{1}{n-1} \cdot \sum_{i=1}^{n} \left(\left(\widetilde{x}_i^{(2)} - \widetilde{x}_i^{(1)} \right) - m \right)^2.$$

From $V = 2\sigma^2$, we can reconstruct σ as $\sigma = \sqrt{V/2}$.

Problem. For eddy covariance towers, the tower is the best available measuring instrument, so no more accurate ("standard") measuring instrument is available. Thus, the first way of gauging random error is not applicable in our case.

The second way is, in principle, possible: we can bring all the instruments from one tower to a location of another tower and thus, compare the measurement results. However, while this is possible (and done), this is a very expensive and time-consuming procedure, so it can be used only rarely.

It is therefore desirable to come up with an easier way of gauging the random error of eddy covariance measurements.

What we do in this paper. In this paper we propose a new method for estimating the random error of eddy covariance measurements.

2 Main Idea

The objective of the tower is to measure the flux over a big area and average the values over a half-hour period. This flux is smoothly changing during the day. For example, over plant-covered area, the daily changes in the CO_2 flux reflect the plant biology. During daytime, the plants' photosynthesis process breaks down the CO_2 so its concentration decreases resulting in a downward flux. During nighttime, the plants breathe and produce CO_2, so we have an upward flux. As a result the actual downward flux smoothly increases starting from early morning and then smoothly decreases starting with late afternoon.

While the actual flux smoothly changes with time, the measurement results are non-smooth or "wiggly". This non-smoothness is caused by the difference between the measurement results and the actual values of the flux, i.e., the measurement errors. To be more precise, the non-smoothness is caused by the random component of the measurement error, because the systematic component is the same for all the measurements and therefore does not introduce non-smoothness. It is therefore reasonable to use the observed non-smoothness to estimate the random component of the measurement error.

3 Towards an Algorithm

We know that the flux usually changes smoothly during the day, i.e., that the differences $f_{d,t+1} - f_{d,t}$ between the actual fluxes $f_{d,t}$ and $f_{d,t+1}$ at two consequent moments of time t and $t+1$ are small. We also know that if we perform the measurements during the same part of the season, then the corresponding differences $f_{d,t+1} - f_{d,t}$ will not change much from one day d to another. As a result, if we had no measurement errors, then, for each moment of time t, the differences between the measured flux values $\widetilde{f}_{d,t+1} - \widetilde{f}_{d,t}$ will be practically the same for all days d.

In reality, the differences between the measured fluxes change from one day to another. As we have mentioned, the reason why the observed flux values behave differently from the actual fluxes is that the observed flux values $\widetilde{f}_{d,t}$ also contain systematic and random measurement errors:

$$\widetilde{f}_{d,t} = f_{d,t} + \Delta f_{{\rm s},d,t} + \Delta f_{{\rm r},d,t}.$$

Because of this, each observed difference $\widetilde{f}_{d,t+1} - \widetilde{f}_{d,t}$ consists of three components:

- the actual difference $f_{d,t+1} - f_{d,t}$ between the fluxes – which doesn't change much from day to day,
- the difference $\Delta f_{\mathrm{s},d,t+1} - \Delta f_{\mathrm{s},d,t}$ between systematic errors – which also does change from day to day, and
- the difference between random errors $\Delta f_{\mathrm{r},d,t+1} - \Delta f_{\mathrm{r},d,t}$, which varies from day to day (and is 0 on average).

$$\widetilde{f}_{d,t+1} - \widetilde{f}_{d,t} = (f_{d,t+1} - f_{d,t}) + (\Delta f_{\mathrm{s},d,t+1} - \Delta f_{\mathrm{s},d,t}) + (\Delta f_{\mathrm{r},d,t+1} - \Delta f_{\mathrm{r},d,t}). \quad (1)$$

The first and the second components do not change from day to day, and the mean value of the third component is 0; thus, the mean value of the sum (1) is equal to the sum of the the first and the second components:

$$m_t \stackrel{\text{def}}{=} E\left[\widetilde{f}_{d,t+1} - \widetilde{f}_{d,t}\right] = (f_{d,t+1} - f_{d,t}) + (\Delta f_{\mathrm{s},d,t+1} - \Delta f_{\mathrm{s},d,t}).$$

So, by estimating the mean difference $\widetilde{f}_{d,t+1} - \widetilde{f}_{d,t}$, we thus get an estimate for the sum of the first and the second components:

$$m_t = (f_{d,t+1} - f_{d,t}) + (\Delta f_{\mathrm{s},d,t+1} - \Delta f_{\mathrm{s},d,t}) = E\left[\widetilde{f}_{d,t+1} - \widetilde{f}_{d,t}\right] \approx \frac{1}{D} \cdot \sum_{d=1}^{D} (\widetilde{f}_{d,t+1} - \widetilde{f}_{d,t}), \quad (2)$$

where D denotes the total number of observation days within a given season.

Now, by taking the difference between the observed differences $\widetilde{f}_{d,t+1} - \widetilde{f}_{d,t}$ and the estimated mean m_t, we get the differences

$$\Delta f_{\mathrm{r},d,t+1} - \Delta f_{\mathrm{r},d,t} = \left(\widetilde{f}_{d,t+1} - \widetilde{f}_{d,t}\right) - m_t$$

between the random errors. Similarly to the above description, it is reasonable to assume that these errors are independent, and thus, to conclude that the variance of these differences is equal to $2\sigma_t^2$:

$$V_t = 2\sigma_i^2 \approx \frac{1}{D-1} \cdot \sum_{d=1}^{D} (\Delta f_{\mathrm{r},d,t+1} - \Delta f_{\mathrm{r},d,t})^2 = \frac{1}{D-1} \cdot \sum_{d=1}^{D} \left(\left(\widetilde{f}_{d,t+1} - \widetilde{f}_{d,t}\right) - m_t\right)^2.$$

Thus, we arrive at the following algorithm for gauging the random error.

4 Resulting Algorithm for Estimating Standard Deviation of Random Error

Input data. As the input, we take flux values $\widetilde{f}_{d,t}$ measured on different days $d = 1, \ldots, D$ at different times $t = 1, \ldots, T$. Usually, measurements are performed every half-hour, so $T = 48$ measurement are performed during the 24-hour day.

Objective. The objective is to compute, for each time t, the standard deviation σ_t which characterizes the random error related to measuring the flux at time t.

Algorithm. First, we compute the differences between the results measured at consecutive times of the same day:

$$\widetilde{f}_{d,t+1} - \widetilde{f}_{d,t}.$$

After that, we compute the means

$$m_t := \frac{1}{D} \cdot \sum_{d=1}^{D} \left(\widetilde{f}_{d,t+1} - \widetilde{f}_{d,t}\right).$$

Then, we compute

$$V_t := \frac{1}{D-1} \cdot \sum_{d=1}^{D} \left(\left(\widetilde{f}_{d,t+1} - \widetilde{f}_{d,t} \right) - m_t \right)^2 .$$

Based on V_t, we estimate the desired standard deviation as

$$\sigma_t = \sqrt{\frac{V_t}{2}}.$$

Testing. We tested this algorithm on the simulated data. We started with a simple quadratic heat flux model where at time $t = 0$ flux $f = 0$, at time $t = 24$ flux $f = 1$, and at time $t = 48$ flux $f = 0$. These three requirements uniquely determine the corresponding quadratic dependence as:

$$f_{d,t} = \frac{(t/2) \cdot (24 - (t/2))}{12^2}.$$

To simulate random measurement errors, we selected $\sigma = 0.1$ and then, for each d and t, we added, to the actual flux values, the results $\xi_{d,t}$ of simulating a normally distributed normal variable with 0 mean and standard deviation σ: $\widetilde{f}_{d,t} = f_{d,t} + \xi_{d,t}$. For thus simulated data, the above algorithm enables to recover the value σ with a reasonable accuracy.

Acknowledgments

This work was supported in part by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence).

References

[1] Aubinet, M., T. Vesala, and D. Papale, *Eddy Covariance: A Practical Guide to Measurement and Data Analysis*, Springer, Dordrecht, Hiedelberg, London, New York, 2012.

[2] Rabinovich, S., *Measurement Errors and Uncertainties: Theory and Practice*, Springer-Verlag, New York, 2005.