THE CHARACTERISTICS OF FRACTIONAL UNCERTAINTY ON EDDY COVARIANCE MEASUREMENT AND ITS APPLICATION TO DATA QUALITY CONTROL

DAISUKE KOMORI(IIS, UNIV. TOKYO), WONSIK KIM (NATIONAL INSTITUTE FOR AGRO-ENVIRONMENTAL SCIENCES), JAEIL CHO (KYUSHU UNIV.)

Daisuke KOMORI

INTRODUCTION

In statistical analysis of sampling data, the uncertainty coming from fluctuations during the measurement of a quantity can be summarized by a sampling error δ. According to the error sources, δ is decomposed into (Bevington and Robinson, 2003)

(1) Random error δ_r caused by instrumental uncertainties or statistical fluctuations (i.e., relevant to sensor tolerance and sampling size)
(2) Systematic error δ_s by faulty equipment calibrations or discrepancies among different observers (i.e., senor calibration and frequency loss)
(3) Illegitimate error δ_i by mistakes or measurement blunders (i.e., surface heterogeneity and atmospheric stationarity)

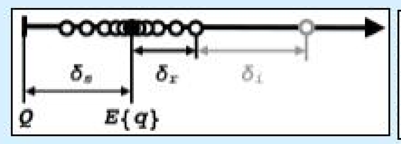


Figure 1: Schematic illustration of error classification on a linear coordinate system. The origin is the true value Q and each cycle is individual measurement q by repeated same experiment to determine Q. The filled black is an expected value E{q} and the opened gray is an outlier estimated by robust regression of q. Each δ_s, δ_r and δ_i represents systematic, random and illegitimate error, respectively.

Nevertheless, several studies on δ analysis to eddy covariance (EC) measurement have merely focused on δ_r and δ_s, and even included δ_i into δ_r or δ_s, although δ_i resulted from not a statistical noise or bias but just an outlier and δ_i was an inevitable consequence over EC measurement owing to use simplified EC equation ignored the terms of advection and horizontal divergence over postulation of surface homogeneity and atmospheric stationarity. Meanwhile, Vickers and Mahrt (1997) presented that mesoscale variability and inhomogeneity were regarded as nonstationarity and it was another constituent of δ along with δ_r and δ_s. In addition, Mahrt (1998) suggested that a method was needed to quantify the nonstationarity of time series, and it should be estimated as part of δ beyond the traditional methods merely considering δ_r and δ_s on account of every atmospheric flow being unable to comply with stationarity. **Understandings so far, δ associated with nonstationarity is assigned to δ_i and it is distinguished from δ_r and δ_s in this investigation.**

Finkelstein and Sims (2001) described a statistically rigorous method to estimate δ as a fractional flux sampling error (i.e. fractional uncertainty ϕ in this study), and suggested that their method was more reliable than previous ones (Wyngaard, 1973; Mann and Lenschow, 1994) in terms of incorporating auto- and cross-covariance terms instead of any preconditional assumptions about time series, boundary layer states, and an autocorrelation function for an appropriate integral time scale. Using this computational method, Kim et al. (2008, 2009) suggested that a robust regression (Rousseeuw and Leroy, 2003) should be applied to determine a random fractional uncertainty ϕ_r during some period because of the difficulty of outlier control in ϕ with EC measurement, and indicated that ϕ_r would be a constant if just δ_r were member of ϕ. The constancy was implied by various literatures subsequent to Meyers et al. (1998) as δ_r was a increasing linear function of its flux (F) magnitude and those slops showed same minimum (Finkelstein and Sims, 2001; Hollinger and Richardson, 2005; Mano et al., 2007; Kim et al., 2008).

In this study, **we investigated the spatiotemporal characteristics of ϕ to demonstrate the constancy of ϕ_r and understand the fundamentals of ϕ_i with EC measurement of various vegetated sites.** It is notable challenges in terms of both uncertainty analysis and quality control and quality assurance (QCQA) issues, because the subject bears upon studies based on worldwide EC measurement with wide applications in ecosystem intercomparison, model development, and model-data synthesis.

MATERIAL AND METHOD

[Fractional uncertainty]

Let q be the value estimated by measuring a quantity Q. Then, the expected value E{q} can be defined by repeated measurements q_t (t = 1, 2, ..., N) in the same experiment as

$$E\{q\} = \lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} q_t$$

Referring to Figure 1, δ_s, which always occurs with the same magnitude in the same direction when measurements are repeated under identical circumstances, and δ_r, which is not dependent of δ_s but occurs in either direction with a different magnitude for each measurement, are defined respectively (Bendat and Piersol, 2000) as

$$\delta_s = E\{q\} - Q$$

$$\delta_r = \sqrt{\mathrm{Var}\{q\}}$$

Supposing δ_s = 0, then, the fractional uncertainty ϕ becomes

$$\phi = \frac{\delta}{|F|} \Rightarrow \phi = \frac{\sqrt{\mathrm{Var}\{\gamma_{xy}\}}}{|\mathrm{Cov}\{x,y\}|}$$

$$\mathrm{Var}\{\gamma_{xy}\} = \frac{1}{n}\left[\sum_{p=-m}^{m}\gamma_{xx}(p)\gamma_{yy}(p) + \sum_{p=-m}^{m}\gamma_{xy}(p)\gamma_{yx}(p)\right]$$

$$\mathrm{Cov}\{x,y\} = \frac{1}{n}\sum_{t=1}^{n}(x_t-\bar{x})(y_t-\bar{y})$$

$$\gamma_{xx}(h) = \gamma_{xx}(-h) = \frac{1}{n}\sum_{t=1}^{n-h}(x_t-\bar{x})(x_{t+h}-\bar{x})$$

$$\gamma_{xy}(h) = \gamma_{yx}(-h) = \frac{1}{n}\sum_{t=1}^{n-h}(x_t-\bar{x})(y_{t+h}-\bar{y})$$

where
- *x* is a vector quantity as vertical wind velocity
- *y* is a scalar such as temperature or mixing ratio of vapor or trace gases
- the symbols ¯ represents average
- *n* denotes the number of samples in an averaging interval τ (i.e., *n* = 36000 in case of τ = 1 h of 10 Hz data)
- *m* is a number of samples sufficiently large to capture the integral timescale (i.e., *m* = 200 because estimated ϕ is unaffected by m > 200 throughout the testing with all our data sets)
- γ_{xx} and γ_{xy} are estimated by auto-covariance and cross covariance of lag *h*

The least-squares method yields the arithmetic mean of the observations which has a Gaussian distribution; however, it might be feasible to estimate the descriptive statistical location in case of measurements including outliers. **Therefore, E{q} needs to be estimated by the least median of squares (LMS) method**, which is a robust statistical method for defining a descriptive statistic location such as

$$E\{q\} = \underset{\hat{\vartheta}}{\mathrm{Minimize}}\ \underset{t}{\mathrm{median}}\ r_t^2$$

where
- ϑ is the regression coefficient of an estimated linear model
- *r* is the residual between the predicted and the estimated dependent variables.

This LMS method was useful to estimate E{ϕ} (henceforth Φ) with F which is independent variable, because various reports has indicated that estimated Φ with EC method consists of not only δ_r but also δ_i in the view point of higher skewness than Gaussian on ϕ distribution (Meyers et al., 1998; Finkelstein and Sims, 2001; Kim et al., 2008, 2009). Its standard deviation s for filtering out δ_i is calculated by

$$s = C\sqrt{\mathrm{median}\ d_t^2}$$

where
- *C* is a factor used to achieve consistency
- $d_t = \phi_t - \Phi$ as residual

[Flux measurement]

Flux measurements for an analysis of Φ were obtained from six sites:

All data were produced by an identical EC system consisting of a three-dimensional sonic anemometer (CSAT3; Campbell Scientific, Utah, USA) and an open-path CO_2/H_2O gas analyzer (LI7500; LI-COR, Nebraska, USA) deployed at various heights and periods among the sites. The hourly flux was calculated by using the methodology of Leuning (2004).

Vegetation type		Site ID	Location			Measurement/ canopy height (m)	Annual Temperature/ Precipitation (C°/ mmy⁻¹)	Available data period
			Lat.	Long.	a.s.l. (m)			
Crop	Cassava	CTT	16°54'	99°25'	156	3/1	22/1100	4-27 Oct 2007
	Paddy	PST	17°04'	99°42'	51	7/1	26/1000	Jul 2005-Sep 2007
	Sugarcane	SPT	16°42'	100°00'	141	6/3	23/1300	9-15 Nov 2007
Forest	Monsoon	DTT11, 13	16°53'	99°26'	114	100, 30/7	22/1100	Jan 2005-Dec 2007
		OTT	16°54'	99°23'	143	10/6	22/1100	3-6 Jun 2007
	Temperate	DDK	37°40'	128°45'	983	30/15	6/1700	Jan-Dec 2004

RESULTS AND DISCUSSIONS

[Short-term analysis]

♦Φ estimated with the LMS method to filter outliers contaminated by δ_i has uniformity over a homogeneous land cover (Figure 2).

The results of earlier studies (Lenschow et al., 1994; Finkelstein and Sims, 2001; Hollinger and Richardson, 2005; Richardson et al., 2008; Vickers et al., 2009) suggested the uniformity, but persuasive data were rare, because the data used in those studies were not free of outliers.

♦An heterogeneous effect on ϕ when windward was along around the border line between different growing conditions due to dramatical changing of wind direction for a week (Figure 3).

(1) from one homogeneous paddy field with leaf area index (LAI) ≈ 0.5 for 24-26 May
(2) along around the border line for 27-28 May
(3) from the other homogeneous paddy field with LAI ≈ 2.0 for 29-30 May

We could deduce that
(1) the magnitude of minimum fractional uncertainty ϕ_{min} was ≈ 0.07 and a constant
(2) estimated Φ reached the constant over homogeneous land cover and well mixed turbulence condition even though heterogeneous one.

♦ϕ_i was mostly distributed around close to zero F, although some ϕ_i was also scattered over the other F magnitudes even in induced turbulent conditions (Figure 2 and Figure 4).

Stationarity issue is cause of δ_i and it could be possible to define ϕ as an indicator of appropriate conditions for EC measurement.

[Long-term analysis]

♦The Φ trends showed little seasonal variability regardless of the type of flux (Figure 5).

The precise of Φ was higher when those magnitude was smaller and the long-term Φ measured heterogeneous site was somewhat larger than homogeneous one.

♦The Φ_H, Φ_{lE}, and Φ_{Fc} values are nearly same among homogeneous sites, but whose increasing trend according to z is captured. It is more clear when heterogeneous sites are included (Figure 6).

On the base of z domain, it is possible to suppose that the magnitude of F are all the same in turbulence layer. Φ uniformity will be reasonable in the layer regardless of z. Because the variability of vertical wind velocity is increased and that of every target scalar quantity is decreased along increasing z. Therefore, variance parameter which is δ in our study is unchangeable according to z, because δ is computed with those two values which have inverse relationship between themselves. However, our result satisfied above hypothesis in sufficient.

It is possible to conclude that
(1) ϕ_{min} is a constant regardless of spatiotemporal variabilities
(2) heterogeneity is reflected in Φ: increasing z might lead to larger Φ
(3) Φ_{lE} and Φ_{Fc} were larger than Φ_H with increasing z, because they have a greater potential for inhomogeneity (Katul et al., 1999; Oren et al., 2006).

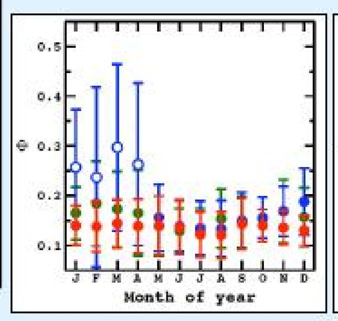


Figure 5: Temporal trend of monthly means of Φ measured at DTT13 from January 2005 to December 2007. Red, blue, and green denote respectively expected Φ_H, Φ_{lE}, and Φ_{Fc}, and the tails represent the standard deviation of Φ with the least median of squares method.

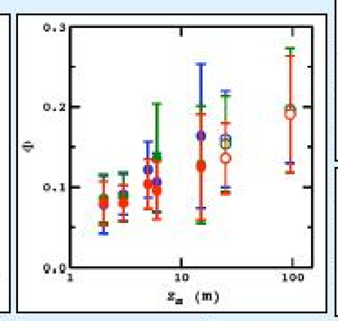
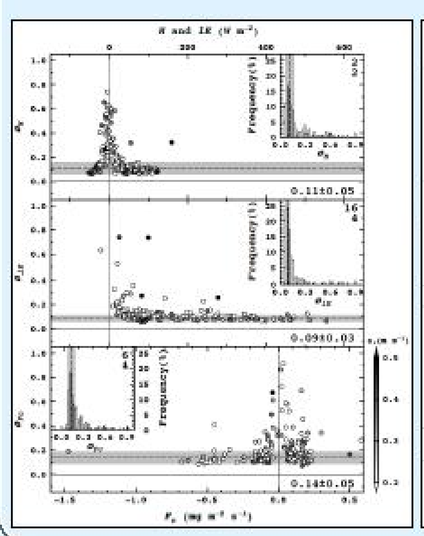
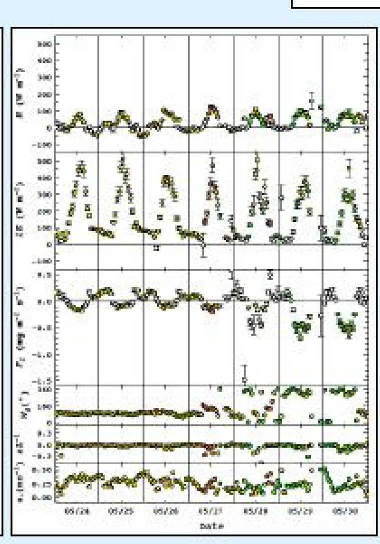


Figure 6: Variability of Φ according to flux measurement height from mean vegetation height z_m over all vegetation types. Symbol are the same as Figure 5, and the open circles denote the Φ at heterogeneous sites (DTT11 and DTT13).





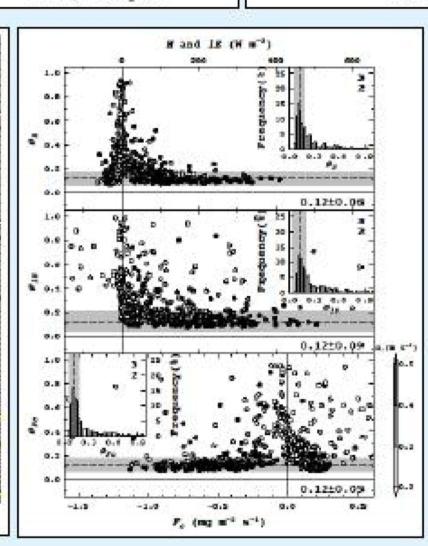


Figure 2: Relations between ϕ and the flux values for sensible heat, latent heat and CO_2 (H, lE and F_c), and the frequency distribution of ϕ over PST during 24–30 May 2007. Dashed lines denote the expected Φ and gray regions show the range of one standard deviations s by the least median of squares method in every panel. The inserted numbers at right-bottom in each relationship panel are Φ ± s, and at right-top in each frequency distribution panel denote kurtosis and skewness, respectively. The brightness of the symbols represents the scale of the friction velocity u_*.

Figure 3: Time series of H, lE, and F_c, and micrometeorological variables (wind direction Wd; atmospheric stability zL^{-1}, where z is measurement height and L is the scale of the Monin-Obukhov length; friction velocity u_*) at PST during 24–30 May, 2007. Symbols are colored in accordance with Wd which is the azimuth measured clockwise from the head of the ultra sonic anemometer (yellow: 110°, green: 360), and represent accepted measurements based on the expected fractional uncertainty Φ (0.1±0.05) criterion; vertical tails denote sampling error δ.

Figure 4: Same as Figure 2 except measured over DTT13 on August 2005.

[Averaging timescale analysis]

Releasing the flux averaging interval by 1 h and supposing Φ arises entirely from δ_r, Φ decreases with an increase in the averaging scale τ and closes to zero as τ approaches infinity over theoretical base (Moncrieff et al., 1996; Bendat and Piersol, 2000; Katul et al., 2004; Vickers et al., 2009).
Consequntly, Φ as a function of τ can be defined as

$$\Phi = \rho\tau^{-\frac{1}{2}},$$

where ρ denotes Φ at unit τ (i.e., 1 h). Considering Φ uniformity estimated on finest δ_r and then taking the measured quantity 0.08 ± 0.03 derived from CTT and SPT as the constant fractional uncertainty ω to substitute for ρ, we can in final define a baseline function of fractional uncertainty Φ_{base} as theoretical base:

$$\Phi_{base} = \omega\tau^{-\frac{1}{2}}$$

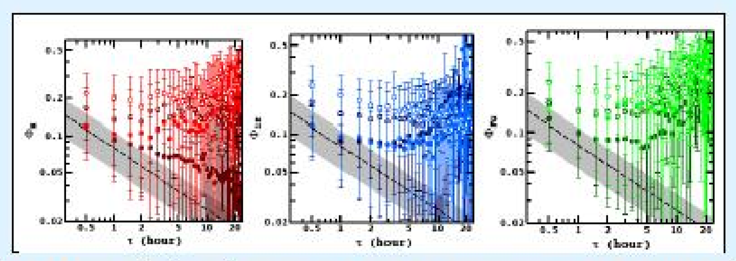


Figure 7: Expected Φ_H, Φ_{lE}, and Φ_{Fc} by averaging timescale τ at various measurement sites (DTT11: open bright circles; DTT13: open dark, circles; PST: closed bright circles; and CTT: closed dark circles) for approximately one month of data during the rainy season. Tails of the symbols represent the standard deviation of Φ, and dashed line and gray region describe $\Phi_{base} = \omega\tau^{-1/2}$ with those territory of standard deviation. The ω, standing for constant Φ at unit τ (i.e., 1 h), is 0.08 estimated by CTT and SPT data sets.

♦The Φ values from sites CTT and PST matched the Φ_{base} line at τ = 0.5 hour, and were parallel to the initial slopes defined by measurements with τ = 0.5 and 1 hour.; however, most of the other regions were different from Φ_{base}.

These results suggest that effects of land surface heterogeneity at measurement sites and diurnal cycles in variables relevant for flux estimation on Φ was well demonstrated.

Considering that the heterogeneity of EC measurements is dependent upon τ, since higher values of τ diminish the spatial variability of single-tower EC measurements (Raupach and Shaw, 1982; Katul et al., 2004; Vickers et al., 2009), our data support the findings of previous investigations at least τ ≤ 1 h. Therefore, the heterogeneity scale parameter η can be developed as

$$\eta = 1 - \frac{\omega}{\rho}$$

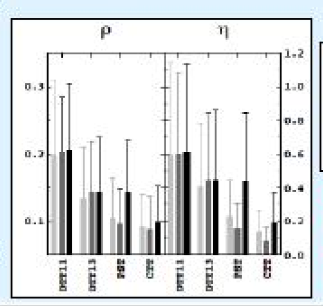


Figure 8: Expected fractional uncertainty ρ and estimated heterogeneity scale η of 1 h averaging scale according to fluxes and sites using data of Figure 7. Brightness and site abbreviation are the same as in Table 1 and Figure 7, respectively. Tails denote the error values calculated with error propagation formulas using the standard deviations of Φ.

Our results demonstrate that a visualized homogeneous landscape does not always satisfy a homogeneous condition for EC flux measurements.

CONCLUSION

We estimated the expected fractional uncertainty Φ over various types of land cover with least median of squares (LMS) method (Rousseeuw and Leroy, 2003) to understand the characteristic of fractional uncertainty ϕ on eddy covariance (EC) flux measurement. Estimated Φ over homogeneous land cover closed to a constant, and even though minimum fractional uncertainty ϕmin over heterogenous land cover approched the constant in irrespective of temporal scale and kind of flux. The theoretical function according to averaging scale τ was defined as $\Phi = \omega\tau^{-1/2}$, where ω denoted the constant of fractional uncertainty, and estimated as 0.08 ± 0.03 at τ = 1 h based on EC measurement. Because estimated Φ for heterogeneous land cover larger than ω defined at homogeneous sites, we suggest the use of a heterogeneity scale parameter η = 1 − (ω/ρ), where ρ is Φ in this study. These results suggest that ω can be a general threshold to support data quality control and assurance (QCQA), ϕ analysis is adequate to estimate not only uncertainty magnitude but also whose sources, and Φ analysis is helpful to explore ecosystem intercomparison, model development, and model-data synthesis with EC measurement.