Science in China Series D: Earth Sciences

www.scichina.com
earth.scichina.com
www.springerlink.com

Uncertainty analysis of CO_2 flux components in subtropical evergreen coniferous plantation

LIU Min[1,2,3], HE HongLin[1†], YU GuiRui[1], LUO YiQi[4], SUN XiaoMin[1] & WANG HuiMin[1]

[1] Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China;
[2] Gradute School of the Chinese Academy of Sciences, Beijing 100049, China;
[3] School of Geography Science, Nanjing Normal University, Nanjing 210097, China;
[4] Department of Botany and Microbiology, University of Oklahoma, 770 Van Vleet Oval, Norman, OK 73019, USA

We present an uncertainty analysis of ecological process parameters and CO_2 flux components (R_{eco}, *NEE* and gross ecosystem exchange (*GEE*)) derived from 3 years' continuous eddy covariance measurements of CO_2 fluxes at subtropical evergreen coniferous plantation, Qianyanzhou of ChinaFlux. Daily-differencing approach was used to analyze the random error of CO_2 fluxes measurements and bootstrapping method was used to quantify the uncertainties of three CO_2 flux components. In addition, we evaluated different models and optimization methods in influencing estimation of key parameters and CO_2 flux components. The results show that: (1) Random flux error more closely follows a double-exponential (Laplace), rather than a normal (Gaussian) distribution. (2) Different optimization methods result in different estimates of model parameters. Uncertainties of parameters estimated by the maximum likelihood estimation (MLE) are lower than those derived from ordinary least square method (OLS). (3) The differences between simulated R_{eco}, *NEE* and *GEE* derived from MLE and those derived from OLS are 12.18% (176 g $C \cdot m^{-2} \cdot a^{-1}$), 34.33% (79 g $C \cdot m^{-2} \cdot a^{-1}$) and 5.4% (92 g $C \cdot m^{-2} \cdot a^{-1}$). However, for a given parameter optimization method, a temperature-dependent model (T_model) and the models derived from a temperature and water-dependent model (TW_model) are 1.31% (17.8 g $C \cdot m^{-2} \cdot a^{-1}$), 2.1% (5.7 g $C \cdot m^{-2} \cdot a^{-1}$), and 0.26% (4.3 g $C \cdot m^{-2} \cdot a^{-1}$), respectively, which suggested that the optimization methods are more important than the ecological models in influencing uncertainty in estimated carbon fluxes. (4) The relative uncertainty of CO_2 flux derived from OLS is higher than that from MLE, and the uncertainty is related to timescale, that is, the larger the timescale, the smaller the uncertainty. The relative uncertainties of R_{eco}, *NEE* and *GEE* are 4%—8%, 7%—22% and 2%—4% respectively at annual timescale.

CO_2 flux components, statistical uncertainty analysis, bootstrapping method, subtropical evergreen coniferous plantation, Qianyanzhou

Ecosystem carbon budget is one of the important scientific problems in global change research[1]. It is significant to study CO_2 carbon exchange between terrestrial ecosystem and atmosphere for understanding the carbon budget in regional and global scales. Eddy covariance method is the most direct method at present for flux measurement in CO_2 exchange research. On the one hand, net ecosystem exchange (*NEE*) observed by this method is assimilated by ecological models. Data assimilation methods are also used in filling flux data and separating flux components[2,3]. Generally speaking, *NEE* is the balance between gross ecosystem exchange (*GEE*) and ecosystem respiration (R_{eco})[4]. In order to

Received January 1, 2008; accepted May 15, 2008; published online January 5, 2009
doi: 10.1007/s11430-009-0010-6
†Corresponding author (email: hhonglin@cern.ac.cn)
Supported by National Natural Science Foundation of China (Grant No. 30570347), Innovative Research International Partnership Project of the Chinese Academy of Sciences (Grant No. CXTD-Z2005-1) and National Basic Research Program of China (Grant No. 2002CB412502)

forecast the change of ecosystem carbon budget, simulation and validation of ecosystem processes have to be done through full uses of CO_2 flux observation, CO_2 flux components and ecological models[5,6]. However, due to the deficiency in measurement and simulation technology, there is a great uncertainty in simulating CO_2 exchange between biosphere and atmosphere. With the development of model-data fusion research in terrestrial ecosystem, flux data uncertainty is equally important as data themselves. How to quantify the uncertainty of flux data, the uncertainty of key parameters in ecological process and different flux components has become a frontier issue and a focus of global flux research, and it also is an important part in global carbon cycle study[7].

At present, much research on flux data uncertainty has been done internationally. Hollinger et al.[6] have studied flux data uncertainty through using repeated sampling method in two nearby towers in Howland Forest, pointing out that flux measurement error follows a double-exponential (Laplace) distribution rather than a normal (or Gaussian) distribution. Meanwhile, a daily-differencing approach was proposed to the random errors of flux measurements in single tower, which calculates differences of measured *NEE* between two points of time in adjacent days with similar climate conditions to estimates of measurement errors. The effect of different error distributions on *NEE* and model parameter was also considered. Richardson et al.[7] compared the effects of two different error distributions (Gaussian distribution and double-exponential distribution) and two parameter optimization methods on model parameters and the respiration component, demonstrating that the uncertainty from parameter optimization methods is larger than that from different models, and they also pointed out that it may be controversial to use a maximum likelihood method as the parameter optimization method, which may not be appropriate in all situations. Richardson et al.[8] conducted a cross-site study on flux measurement errors in AmeriFlux, including forest, agricultural and grassland ecosystems, and demonstrated that flux measurement errors of different ecosystems follow a double-exponential distribution. The relationships between measurement error and environment variables (wind speed, *PPFD*, etc.) and flux magnitudes were also examined. Papale et al.[9] analyzed flux observation in EUROFLUX at 8 sites during 12 years, concluding that all data processing methods had important influence on *NEE* and its uncertainty. Falge et al.[10] studied the influences of different gap-filling methods (mean diurnal variation method, nonlinear regression method and look-up table method) on annual total *NEE* and its uncertainty. They found that there was not obvious difference between different methods. Hui et al.[11] filled the data gap of *NEE*, latent heat and sensible heat at 3 sites of AmeriFlux and computed the confidence intervals of these 3 fluxes. Hagen et al.[12] presented uncertainty analysis of *GEE* estimates derived from 7 years' continuous eddy covariance measurements in Howland Forest, considering the difference between simple ecological models and artificial neural network (ANN) model, other than the difference between different ecological models.

In China, the establishment of ChinaFLUX provides a platform for research on CO_2, H_2O and energy exchange between China terrestrial ecosystems and atmosphere. However, no systematic uncertainty analysis has been conducted on flux observations. How to quantify and reduce the uncertainty of flux data and obtain representative data is an urgent task for ChinaFLUX. Those also are important scientific problems widely concerned in international flux research networks. In this paper, we present an uncertainty analysis of ecological process parameters and CO_2 flux components (R_{eco}, *NEE* and *GEE*) derived from 3 years' continuous eddy covariance measurements of CO_2 fluxes in subtropical evergreen coniferous plantation, Qianyanzhou, of ChinaFlux. Random errors of CO_2 fluxes measurements are obtained by the daily-differencing approach. We used an uncertainty analysis method (bootstrapping) to qualify the statistical uncertainty of CO_2 flux components. Variation of flux uncertainty with timescales is also explored. To our knowledge, this is the first study on observations uncertainty in ChinaFLUX, which can provide a technical support for correctly evaluating flux observations, quantifying flux observations uncertainty, establishing evaluation method system for observations uncertainty and constructing carbon cycle model-data fusion system in China.

1 Data

1.1 Site description

Qianyanzhou coniferous flux tower (26°44′29.1″N, 115°03′29.2″E), set up in August 2002, is located in Qianyanzhou Agriculture Experimental Station of Red

Soil and Hilly Land of CERN in Jiangxi Province, South China. The site belongs to the typical subtropical monsoon climate. The coniferous forest plantation was mainly planted in around 1985. The mean canopy height is appropriately 12 m. The forest canopy is dominated by *Pinus massoniana* Lamb, *Pinus elliottii* Engelm, and *Cunninghamia lanceolata* Hook. There also are other plants, such as *Schima crenata* and *Korthals,Cirtus.* According to the statistics of meteorological data from 1985 to 2002, the mean annual temperature is 17.9°C, mean annual precipitation is 1542.4 mm and mean annual evaporation is 1110.3 mm[13].

1.2 Data collecting and processing

Flux data (water, carbon and heat) and conventional metrological measurements during 2003－2005 were used in this study. Flux data were measured by open-path eddy covariance system at 39 m. Liu et al.[13,14] and her colleagues have described the data and instruments in detail. Data have been corrected and selected by ChinaFLUX flux CO_2 data processing system[15] before analyzing, including: (1) coordinate rotation for 30 min flux data, (2) Webb-Pearman- Leuning correction, (3) *NEE* storage calculation, (4) abnormal data rejection, and (5) nighttime filtering with u^* threshold equaling $0.2\ \mathrm{m \cdot s^{-1}}$[13].

2 Method

Flux data uncertainty usually includes uncertainty from measurement (observation uncertainty) and uncertainty from model parameters (model uncertainty)[16]. Measurement uncertainty can be divided into system error and random error. In most cases, system error is usually a whole deviation, which is hard to decide. Compared to system error, random error is the result of incomplete spectral response and inhomogeneous turbulence mix, which is easier to be quantified with statistical value[16]. Therefore, we only analyze random error in this paper. Model uncertainty is often caused by the deficiency of consensus on the best models to be used, or optimization criteria[7]. The uncertainties caused by different error distributions, different models and different parameter optimization methods are demonstrated without considering system error and uncertainty caused by friction velocity. To carry out a systematic CO_2 flux statistical uncertainty study, the following components are needed: (1) random error analysis method for flux data, (2) ecological process model, (3) model parameter optimization methods, and (4) uncertainty analysis method for estimation of parameters and flux components.

2.1 Analysis on random error of flux measurements

The uncertainty in measurements can be defined as the variance of high-frequency data in average time[7] (e.g. 30 min), which can be detected by making multiple measurements when the data are relatively independent and the condition is stable, and then using the variability of these measurements to estimate the standard deviation. However, CO_2 flux is usually not stable with the influence of phenology and climate. Therefore, people attempt to use simultaneous measurements from two towers located nearby to meet the assumption of repeated sampling method [6]. Hollinger et al.[6] used simultaneous measurement from two towers, located 775 m apart, to simulate CO_2 flux measurement errors, and the result is similar to and consistent with that derived from a traditional micrometeorological method.

However, there are few sites where two adjacent towers simultaneously measure fluxes from patches of similar vegetation. Therefore, Hollinger et al.[6] prompted a daily-differencing approach to characterize flux uncertainty in a single tower. The daily-differencing approach generally resulted in estimates of flux uncertainty that were about 25% higher than those obtained by the two-tower method. Because there is only one tower in Qianyanzhou, we use the daily-differencing approach to quantify the flux measurement errors. A measurement pair (x_1, x_2) is considered valid only if both measurements are made under "equivalent" environmental conditions (*PPFD* within 75 $\mu\mathrm{mol \cdot m^{-2} \cdot s^{-1}}$, air temperature within 3℃, and wind speed within 1 m/s) in the same successive two days. We used $(x_1 - x_2)/\sqrt{2}$ to express the measurement errors. Probability distribution function (PDF) and standard deviation of random flux errors are used to characterize flux measurement uncertainty. Finally, we can estimate random flux errors ($\sigma(\delta)$) by calculating the standard deviation of the differences:

$$\sigma(\delta) = \frac{1}{\sqrt{2}} \sigma(X_1 - X_2). \tag{1}$$

2.2 Ecosystem process models

Generally speaking, *GEE* is usually the difference between *NEE* and ecosystem respiration:

$$GEE = NEE - R_{\mathrm{eco}}, \tag{2}$$

where the positive sign in *NEE* shows that the ecosystem releases CO_2 to the atmosphere, and the negative sign represents the ecosystem uptake of CO_2 from the atmosphere. *GEE* and gross ecosystem productivity (*GPP*) are similar in magnitude but opposite in sign, namely $GEE=-GPP$. *GEE* is equal to zero at night as there is no photosynthesis and observed *NEE* equals ecosystem respiration. *GEE* is the difference between *NEE* and R_{eco} during the daytime, which equals the difference between observed *NEE* and simulated R_{eco} when *NEE* is valid. However, if there is a missing in daytime *NEE*, *GEE* is the difference between simulated *NEE* estimated through an *NEE* model and valid observations and simulated R_{eco}.

Temperature and soil water availability are two important environmental variables for regulating ecosystem respiration. As rainfall and heat are not in the same season in Qianyanzhou, water stress will produce influence on growth of forest. Therefore, we choose Lloyd & Tayor (eq. (3)) to represent temperature effects on ecosystem respiration. We also used a Q_{10} model to describe effects of temperature and soil moisture on ecosystem respiration according to a simple Van't Hoff function (eq. (4))[17−19].

$$R_{\mathrm{eco}} = R_{\mathrm{e,ref}} \mathrm{e}^{E_0\left(\frac{1}{T_{\mathrm{ref}}-T_0}-\frac{1}{T_{\mathrm{soil}}-T_0}\right)}, \tag{3}$$

$$R_{\mathrm{eco}} = R_{\mathrm{e,ref}} \mathrm{Q}_{10}(S_{\mathrm{w}})^{\frac{(T_{\mathrm{soil}}-T_{\mathrm{ref}})}{10}}, \quad Q_{10}(S_{\mathrm{w}}) = a + bS_{\mathrm{w}}, \tag{4}$$

where R_{eco} ($\mu mol \cdot m^{-2} \cdot s^{-1}$) is the ecosystem respiration, $R_{e,ref}$ ($\mu mol \cdot m^{-2} \cdot s^{-1}$) is the ecosystem respiration at the reference temperature (T_{ref}), which is set to 15℃ in this study, namely 288.15 K, E_0 (K) is the activation energy and is fixed to 309 K, T_{soil} (℃) is soil temperature, and S_w ($m^3 \cdot m^{-3}$) is soil moisture. We select T_{soil} and S_w as the temperatures and moisture of soil in 5 cm, *a* and *b* are test parameters, and a positive *b* would mean, for example, that the temperature sensitivity of ecosystem respiration increases with increasing soil water content.

The response of *NEE* to photosynthetic photo flux densities (*PPFD*) could be described as a rectangle hyperbola curve known as the Michaelis-Menten model (eq. (5))[20,21]:

$$NEE = \frac{P_{\max} PPFD}{K_{\mathrm{m}} + PPFD} - R, \tag{5}$$

where *PPFD* ($\mu mol \cdot m^{-2} \cdot s^{-1}$) is the photosynthetic photo flux densities, P_{max} ($\mu mol \cdot m^{-2} \cdot s^{-1}$) is the maximum rate of photosynthesis, and K_m is a Michaelis-Menten constant, which equals the photosynthetic half-saturation constant, and *R* ($\mu mol \cdot m^{-2} \cdot s^{-1}$) is the daytime ecosystem respiration estimated by respiration models (eqs. (3) and (4)). Therefore, we used two models to calculate *NEE*.

For simplicity, models (eqs. (2), (3), and (5)) are selected as a group in which ecosystem respiration is simulated according to temperature functions (i.e., T_model) while models (eqs. (2), (4), and (5)) are selected as a group with ecosystem respiration being modeled according to both temperature and moisture functions (i.e., TW_model).

2.3 Parameter optimization methods

Most model parameter optimizations to date have been based on ordinary least square method (OLS), with the optimization criterion of least squares fitting. The method is simple and convenient, even without knowing the error distribution of measurements and model parameters, but the precision is not high. The maximum likelihood method considers the maximally occurring probability, from which maximum likelihood estimation (MLE) of parameters is obtained.

Generally speaking, observation (y_i) is the sum of the "true" value and an error (Δy_i)[22]. The ordinary least squares method yields unbiased parameter estimates when the data meet the assumptions of normality and homoscedasticity[22]. The cost function is based on minimizing the sum squares of deviation between observed and modeled values:

$$F_{\mathrm{C\_LS}} = \sum_{i=1}^{N} \left(\frac{(y_i - y_{\mathrm{pred}})}{\sigma_i} \right)^2. \tag{6}$$

It would be specially mentioned that ordinary least squares estimation is maximum likelihood when the error is normally distributed and homoscedasticity.

However, when these assumptions do not agree with each other, the parameters estimates are not unbiased. It is necessary to find a new optimization method, such as the maximum likelihood method. Especially, given the double-exponential distribution (eq. (7)), the maximum likelihood function is presented as eq. (8).

$$f(x) = \frac{\mathrm{e}^{-\left|\frac{x-\mu}{\beta}\right|}}{2\beta}, \qquad \sigma = \sqrt{2}\beta, \tag{7}$$

where μ is the location parameter, β is the scale parameter and σ is standard deviation.

$$\text{Prob}\{y_i - y_{\text{pred}}\} \sim \exp\left(-\left|\frac{y_i - y_{\text{pred}}}{\sigma_i}\right|\right), \tag{8}$$

where y_{pred} is the prediction value containing parameters information, and we see that in this case, the maximum likelihood estimator is obtained by minimizing the mean absolute deviation rather than the mean square deviation.

Based on absolute deviation criterion[23], the cost function is expressed as

$$F_{\text{C\_AD}} = \sum_{i=1}^{N}\left(\frac{|y_i - y_{\text{pred}}|}{\sigma_i}\right). \tag{9}$$

Obviously, for maximum likelihood optimization, the distribution or standard deviation (σ) of error Δy_i should be known. In most cases, Δy_i is an approximate expressed as model fit residuals or measurement errors. Model fit residuals consist of flux measurement errors and the errors in model fitting process. When we only consider the uncertainty of parameters, flux measurement errors are usually expressed as Δy_i. However, if the flux measurement error is not easy to obtain, or multi-error is comprehensively considered, model fit residuals are used to express Δy_i. In this study, model fit residual (ε) is chosen to express Δy_i.

In order to characterize the ecosystem process as true as possible, MLE and OLS optimization methods are used to estimate parameters, with the assumption that error deviation is unchanged, where parameters of respiration models are fitting at annual timescale, and parameters of daytime *NEE* models are fitting at monthly timescale.

2.4 Uncertainty analysis

Compared with traditional analysis methods applied to data with heteroscedasticity and non-normal distribution, bootstrapping method provides a more reasonable solution to estimate statistical uncertainty[24]. Bootstrapping is essentially a Monte Carlo method. In the bootstrapping procedure, a synthetic data set is generated by randomly sampling, and then the statistics are estimated. Based on the repeated sampling of empirical data and its correlation estimation, bootstrapping can increase the estimation precision of confidence interval and critical value[25].

There are four groups of methods in this research, namely a combination of the TW_model and the OLS method (TW_OLS), a combination of the TW_model and the MLE method (TW_MLE), a combination of the T_model and the OLS method (T_OLS) and a combination of the T_model and the MLE method (T_MLE). We use the bootstrapping method to analyze uncertainty of these four combinations. The relative uncertainty of ecological process parameters and CO_2 flux components are estimated as Figure 1.

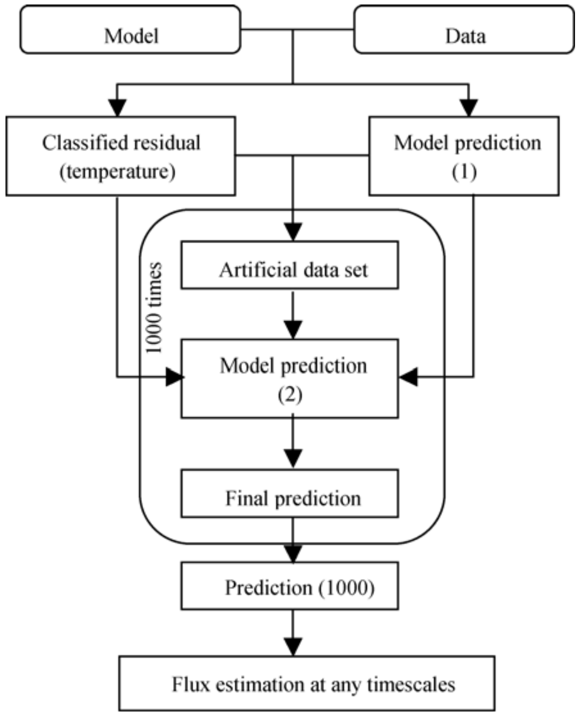

Figure 1 Flowchart of uncertainty analysis.

In step 1, the ecological process models (T_model or TW_model) are fitted to the 30 min valid observations. The residuals from this fit are calculated as the difference of model fitting values and actual observations. The range of environmental variable (e.g. 5 cm soil temperature) is divided into several intervals (we used 10 intervals), and the residuals are binned on the basis of environmental variable value at the time of measurement.

In step 2, an artificial data set is created by adding the "model fit" predicted values to random residuals drawn with replacement from the correct bin (using bootstrapping method).

In step 3, revised ecological models are fit to the bootstrapping data set, and then this model is used to predict flux values for the gap points.

In step 4, repeat steps 2 and 3 more than *N* times (we used *N*=1000). Every 30 min gap point in the time series will have 1000 estimated values, and statistical characteristics (e.g. mean and variance) are calculated using the distribution of the results.

In step 5, 1000 complete component flux time series are generated by using the measured value at every point in the time series where there is an observation and by using a bootstrapping-predicted value for those time steps with no measurement. Sums of fluxes at different timescales are estimated from these N synthetic time series.

In this paper, relative uncertainty (RU) is expressed as the magnitude of the mean 90% prediction interval (i.e., the difference of upper limit and lower limit of 90% prediction interval) divided by the mean prediction value:

$$\mathrm{RU} = \frac{90\% \text{ prediction interval}}{\text{mean prediction value}} \times 100\%. \tag{10}$$

3 Results and discussion

3.1 Random error and model residuals of flux measurements

With the daily-differencing approach, 4170 measurement pairs are derived from 3 years' (2003—2005) continued measurements at Qianyanzhou. We used these pairs to calculate the measurement errors (δ) in measured fluxes. Meanwhile, residuals are also calculated by fitting ecological process model (e.g., T_OLS) to 7549 *NEE* valid observations. The probability distributions of random flux error and model residuals are showed in Figure 2. We find that the double-exponential distribution provides a better fit to the error (measurement error and model residuals) than the normal (Gaussian) distribution, capturing the high peak and thick tail, which is similar to the error distribution in Howland and other towers in AmeriFlux[8]. High peak means that the small error has a higher frequency than the normal distribution, and thick tail means that the big error also has a higher frequency than the normal distribution. These distributions apparently result from several factors. First, for all fluxes, the data are heteroscedastic, with error increasing along with the absolute magnitude of the flux. When this heteroscedastity is combined with the frequency of different flux magnitudes (the instances of low values are far more than high values), a strongly peaked error distribution resulted. The second factor leading to the non-normal error distribution is related to the measurement system. Occasionally, "glitches" caused by power fluctuations, and contamination of other factors result in measured values that are far from correct[6].

The flux measurement error of the double-exponential distribution violates the assumptions for the ordinary least squares fitting with normality and heteroscedastity. Consequently, the OLS optimization method is not suitable for analysis of flux data. We need to introduce a new parameter optimization method (e.g., MLE method) to make a more reasonable analysis for flux data. Meanwhile, the standard deviation of model fit residuals is 4.12 $\mu\mathrm{mol} \cdot \mathrm{m}^{-2} \cdot \mathrm{s}^{-1}$, which is higher than that of random flux error of 3.59 $\mu\mathrm{mol} \cdot \mathrm{m}^{-2} \cdot \mathrm{s}^{-1}$ by 12%. It shows that measurement uncertainty makes an important influence on flux data uncertainty.

3.2 Model parameters and their uncertainties

To obtain model parameters and their uncertainties, we present parameter optimization to the four combinations (TW_OLS, TW_MLE, T_OLS, T_MLE). The parameters of respiration and their uncertainties are showed in

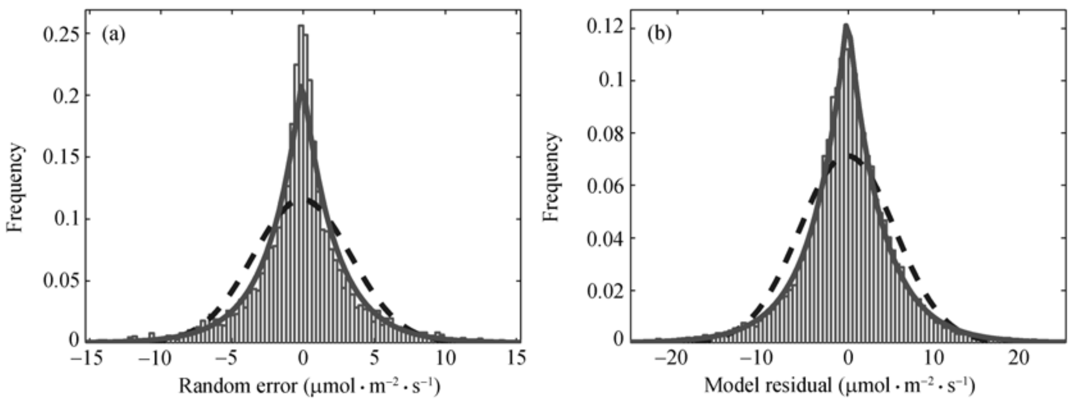

Figure 2 Histogram depicting the frequency distribution of the inferred flux measurement errors and model fit residuals. (a) Flux measurement error; (b) model fitting residuals. The dotted line depicts a normal distribution, whereas the solid line shows a double-exponential distribution.

Table 1. The fitness of the TW_model (considering temperature and moisture) is higher than that of the T_model (only considering temperature). It suggests that soil moisture plays an important role in regulation of ecosystem respiration at Qianyanzhou with a perennial summer drought, especially an abnormal drought in 2003[15], and fitness is significantly increased from 0.40 to 0.46. The reference respiration $R_{e,ref}$ at a given temperature of 15°C estimated with the MLE method is lower than that estimated with the OLS method by about 0.3 $\mu mol \cdot m^{-2} \cdot s^{-1}$. This result is similar to Richardson's result[7] in Howland. Similar results of Q_{10} values in the TW_model are achieved by different model parameter optimizations in different years. The Q_{10} values estimated derived from OLS are slightly higher than those derived from MLE in 2003 and 2004 but the opposite occurs in 2005. However, the difference is very small and ranges from 0.01 to 0.1. Our results differ from Richardson's[7] in Howland. The change of $R_{e,ref}$ and Q_{10} will naturally lead to the change of annual R_{eco}, and the change tendency of annual R_{eco} is to be discussed below.

Photosynthetic parameters, P_{max} and K_m, and their uncertainties are obtained at monthly scale. In 2004, for example, estimated parameter values vary with different respiration models (Table 2). The MLE estimations of model parameters are lower for P_{max} but higher for K_m. For a given optimization method, the difference of P_{max} between the T_model and the TW_model is about 0.3—0.5 $\mu mol \cdot m^{-2} \cdot s^{-1}$. The difference for K_m is about 14—19 $\mu mol \cdot m^{-2} \cdot s^{-1}$ between the models. For a given model, the difference of P_{max} between MLS and OLS optimization methods is 1.5$\mu mol \cdot m^{-2} \cdot s^{-1}$ and 27$\mu mol \cdot m^{-2} \cdot s^{-1}$ for K_m. The results suggest that the parameter optimization methods are more important than the models in determining parameter values.

Bootstrapping parameter estimation ($R_{e,ref}$, T_0, Q_{10}, P_{max}, K_m) is illustrated in Figure 3, which demonstrates the parameter estimation and their distributions. Kolmogorov-Smirnov test of normality is used for all parameters estimation (Table 3). The result shows that these predictions are generally normally distributed. For a given optimization method, parameters uncertainties of different models are similar. Relative uncertainty of the MLE simulations of the same model is obviously lower than that of the OLS method for $R_{e,ref}$ by 0.05%, Q_{10} by

Table 1 Parameter estimation of ecosystem respiration models based on ordinary least square (OLS) and maximum likelihood estimation (MLE) optimization method from 2003 to 2005, and their uncertainties are also demonstrated

Year	Lloyd & Tayor	OLS method	MLE method	Van't Hoff	OLS method	MLE method
2003	$R_{e,ref}$	2.81±0.09	2.51±0.04	$R_{e,ref}$	2.53±0.10	2.24±0.05
	T_0	215.29±2.00	220.98±0.98	a	−0.02±0.27	−0.01±0.18
				b	15.47±2.46	15.51±1.51
				Q_{10}	2.80±0.18	2.78±0.11
	R^2	0.41	0.40	R^2	0.47	0.47
2004	$R_{e,ref}$	2.88±0.11	2.46±0.06	$R_{e,ref}$	2.78±0.12	2.38±0.06
	T_0	229.22±2.04	229.91±1.50	a	1.38±0.40	1.82±0.32
				b	6.39±3.00	3.19±2.29
				Q_{10}	2.28±0.13	2.27±0.08
	R^2	0.50	0.50	R^2	0.51	0.51
2005	$R_{e,ref}$	2.80±0.07	2.40±0.04	$R_{e,ref}$	2.75±0.08	2.36±0.05
	T_0	224.19±1.71	225.87±1.08	a	1.33±0.21	1.18±0.17
				b	5.00±1.54	6.74±1.22
				Q_{10}	2.16±0.09	2.26±0.07
	R^2	0.49	0.49	R^2	0.50	0.49

Table 2 Estimation of Michaelis-Menten model parameters (P_{max}, K_m),and their uncertainties are also demonstrated, where the values illustrated are annual mean values

Parameter	TW_OLS	TW_MLE	T_OLS	T_MLE
P_{max}	25.66±3.53	24.21±2.37	26.04±3.71	24.53±2.45
P_{max}_RU(%)	32	22	31	22
K_m	661.72±195.477	688.52±142.93	680.57±206.64	704.82±146.73
K_m_RU(%)	68	44	67	45

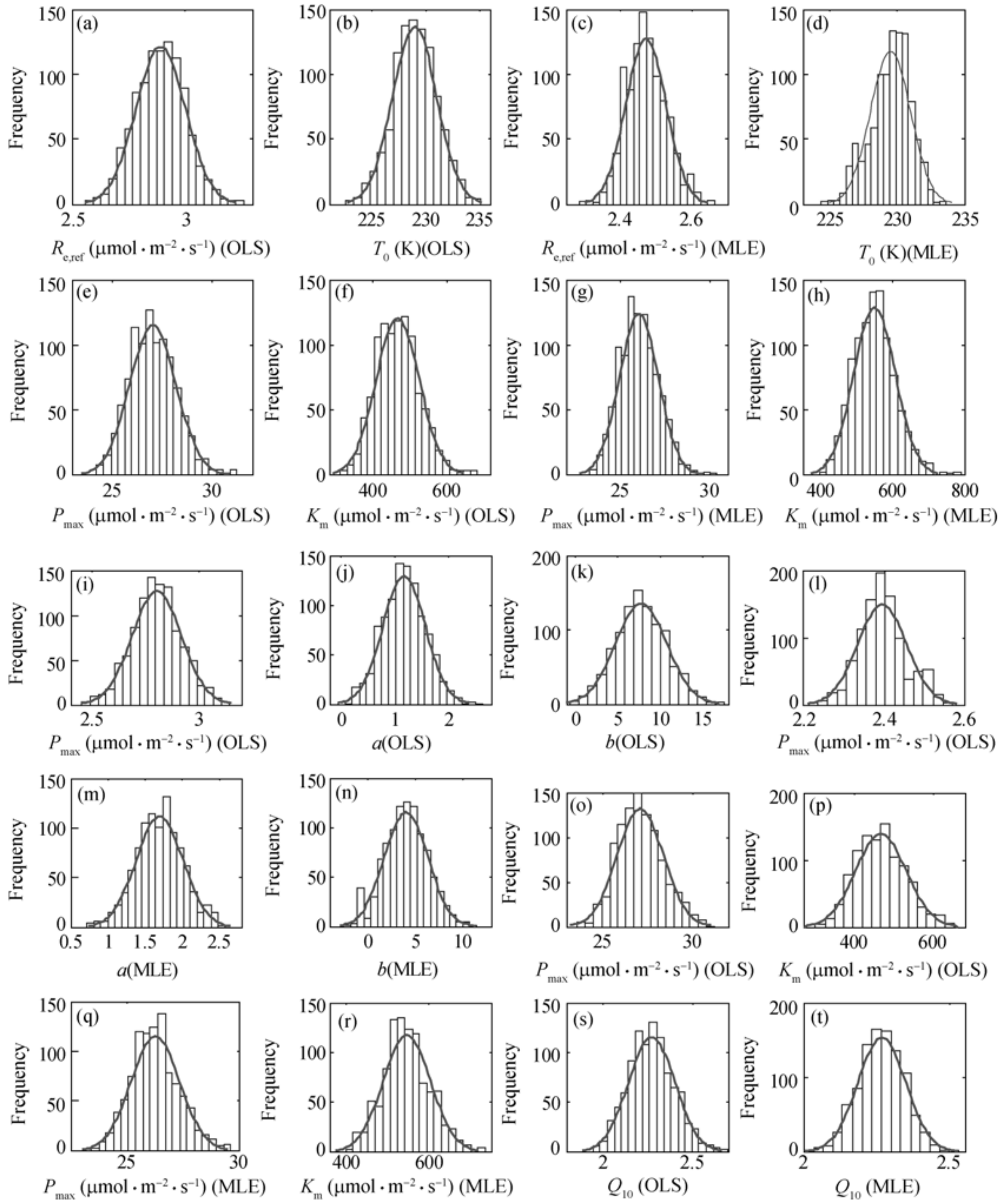

Figure 3 Parameter estimation and their distributions. (a)—(h) Parameter estimation of T_model; (i)—(t) parameter estimation of TW_model, where the content in the brackets means parameter optimization method (i.e. OLS and MLE). The solid line shows a normal distribution whereas respiration parameters ($R_{e,ref}$, T_0, a, b, Q_{10}) are fitted in 2004, and photosynthetic parameters (P_{max}, K_m) are fitted in August, 2004.

Table 3 P values of Kolmogorov-Smirnov test of normality (α = 0.05)

Parameter/TW_model	TW_OLS	TW_MLE	Parameter/T_model	T_OLS	T_MLE
$R_{e,ref}$	0.5435	0.0224	$R_{e,ref}$	0.9044	0.008
a	0.9985	0.9416	T_0	0.8431	0.0001
b	0.7239	0.8021			
Q_{10}	0.5009	0.6497			
P_{max}	0.2925	0.0331	P_{max}	0.2354	0.3976
K_m	0.2276	0.0732	K_m	0.2707	0.549

0.02%, P_{max} by 1.2% and K_m by 55%.

3.3 CO_2 flux components and uncertainty

GEE and its variation are estimated by the simulations of R_{eco} and *NEE*. By applying the bootstrapping algorithm, we generate 1000 values for every 30 min R_{eco}, *NEE* and *GEE* time series. Annual R_{eco}, *NEE* and *GEE* are calculated by summing up the corresponding 30 min time series. Therefore, the uncertainties of these three flux components can be obtained. The simulations and their uncertainties are shown in Table 4.

From 2003 to 2005, annual totals of R_{eco}, *NEE* and *GEE* range from 1214 to 1551 g $C \cdot m^{-2} \cdot a^{-1}$, from −197.9 to −361.3 g $C \cdot m^{-2} \cdot a^{-1}$ and from −1487 to −1813 g $C \cdot m^{-2} \cdot a^{-1}$ respectively, which are consistent with the result of Liu et al.[14]. Flux components in 2004 are higher than those in 2003 and 2005, due to a serious drought in 2003 and many cloudy days with lower *PPFD* in 2005 and with a lower carbon uptake capacity.

With a given respiration model, the MLE estimates of R_{eco} are lower than the OLS estimates from 2003 to 2005. During the three years, the mean differences are 175.1 g $C \cdot m^{-2} \cdot a^{-1}$ based on the TW_model, and 177.8 g $C \cdot m^{-2} \cdot a^{-1}$ based on the T_model, which are consistent with the result of parameters of respiration. For a given model, the mean difference of annual R_{eco} is 9.73 g $C \cdot m^{-2} \cdot a^{-1}$ based on MLE and 25.92 g $C \cdot m^{-2} \cdot a^{-1}$ based on OLS. The result suggests that the parameter optimization methods are much better than the models in influencing estimation of annual R_{eco}. Whether the parameters is obtained by MLE or OLS, the annual R_{eco} estimation based on the T_model is higher than the estimation with the TW_model, which is in contrast with the R_{eco} estimation in 2004 and 2005, and the result suggests that T_model has a overestimating for ecosystem respiration under water scarcity.

On the average, for a given model, the annual *NEE* estimations derived from MLE is about 76.7 g $C \cdot m^{-2} \cdot a^{-1}$ higher than those derived from OLS during 3 years. For a given optimization method, the difference between the simulated *NEE* based on T_model and that based on TW_model is 20 g $C \cdot m^{-2} \cdot a^{-1}$. These results demonstrate that the parameter optimization methods are more important than the models in influencing estimated annual *NEE*.

The annual *GEE* estimations from 2003 to 2005 have a similar tendency to R_{eco}. With a given model, the annual *GEE* derived from MLE are lower 87.8 g $C \cdot m^{-2} \cdot a^{-1}$ with TW_model and 96.1 g $C \cdot m^{-2} \cdot a^{-1}$ with the T_model than those derived from OLS. The mean difference of annual *GEE* estimates between the models is 6.21 g $C \cdot m^{-2} \cdot a^{-1}$ using the MLE method and 12.35 g $C \cdot m^{-2} \cdot a^{-1}$ using the OLS method. Thus, the parameter optimization methods play a more important role in influencing *GEE* estimation than the models. As *GEE* represents the difference of *NEE* and R_{eco}, and the variability caused by respiration parameters and the optimization methods is much higher than the variability of *NEE*, the change directions of *GEE* are similar to R_{eco} with different models or optimization methods.

The different estimations of annual flux components between MLE and OLS optimization methods can be

Table 4 Annual R_{eco}, *NEE* and *GEE* estimation and their uncertainties based on different models and optimization methods

Year	Flux($g\ C \cdot m^{-2} \cdot a^{-1}$)	TW_OLS	TW_MLE	T_OLS	T_MLE
2003	R_{eco}	1391.79±43.29	1245.51±29.32	1427.15±44.31	1252.57±27.05
	RU(%)	6.22	4.71	6.21	4.32
	NEE	−255.03±19.87	−315.54±12.24	−212.50±20.31	−295.91±11.86
	RU(%)	15.58	7.76	19.11	8.02
	GEE	−1646.82±15.66	−1561.05±10.78	−1639.65±15.44	−1548.49±9.65
	RU(%)	3.14	2.27	3.07	2.05
2004	R_{eco}	1550.68±59.43	1331.86±42.22	1520.72±57.24	1319.05±39.34
	RU(%)	7.67	6.34	7.53	5.96
	NEE	−262.33±37.54	−361.26±18.14	−274.49±26.63	−365.94±16.46
	RU(%)	21.00	10.04	19.41	9.00
	GEE	−1813.02±21.06	−1693.12±15.13	−1795.22±19.35	−1684.99±14.04
	RU(%)	3.81	2.87	3.56	2.76
2005	R_{eco}	1383.39±40.20	1223.09±28.60	1370.95±41.63	1213.77±29.25
	RU(%)	5.81	4.68	6.07	4.82
	NEE	−197.85±18.70	−268.50±12.88	−203.51±18.55	−273.65±12.89
	RU(%)	18.90	9.60	18.24	9.42
	GEE	−1581.23±14.17	−1491.59±10.07	−1574.46±14.07	−1487.42±10.14
	RU(%)	2.92	2.19	2.94	2.25

understood as the difference between optimization criterions. In the mathematical sense, the least square criterion of OLS makes outliers (which may have no biological significance) produce a much stronger influence on the figure of merit[6] because the deviations are squared. Consequently, the MLE method with the absolute deviation criterion decreases the influence of outliers on parameter estimation, and reduces the influence of outlier on annual flux components sums.

Compared to the MLE method, OLS has an overestimated R_{eco} and underestimated *NEE*, which is consistent with the result of Richardson et al.[7]. The optimization methods are more important than the models in influencing estimation of flux components. Their effect on *NEE* is expected to be considerably smaller than that on R_{eco} ,and the reason is that annual sum of *NEE* consists of almost all shares of measured and modeled data (e.g., over three years at Qianyanzhou, 70%—80% of all daytime measurement periods in each year had valid measurement), whereas when annual sum of respiration is estimated, roughly 60%—70% of data points (the missing nighttime data plus all daytime respiration) have to be modeled. Therefore, the optimization methods exert more influence on respiration than that on net exchanges. The difference of annual estimations of R_{eco} reached to 175.1 g $C\cdot m^{-2}\cdot a^{-1}$, which is comparable in magnitude to the effect of setting different plausible u* thresholds for nighttime filtering. For example, increasing the u* threshold from 0.1 m/s to 0.25 m/s increases the annual estimated respiration from 105 to 220 g $C\cdot m^{-2}\cdot a^{-1}$ [26]. Thus, enough attention should be attracted in the scientific community.

Bootstrapping algorithm produces empirical probability distributions for each flux component, which is approximately normally distributed (Figure 4, takes only *GEE* for example). The uncertainties of R_{eco}, *NEE* and *GEE* at annual timescale are ±27—60, ±10—40, ±12—25 g $C\cdot m^{-2}\cdot a^{-1}$ respectively, with the relative uncertainties of 4%—10%, 7%—20%, and 2%—3%. *NEE* uncertainty is similar in magnitude reported in the literature (±25 g $C\cdot m^{-2}\cdot a^{-1}$[8]; ±30 g $C\cdot m^{-2}\cdot a^{-1}$[27]; ±20—150 g $C\cdot m^{-2}\cdot a^{-1}$[28]; ±50 g $C\cdot m^{-2}\cdot a^{-1}$[29]; ±40 g $C\cdot m^{-2}\cdot a^{-1}$[30]; –30—+80 g $C\cdot m^{-2}\cdot a^{-1}$[31]). The reason for the difference is probably the inconsistent calculation methods of uncertainty. Uncertainty of R_{eco} is higher than that of *NEE*, suggesting that uncertainty of R_{eco} is the main resource of *GEE* uncertainty. *GEE* and R_{eco} have a greater magnitude than *NEE*, consequently the *NEE* has a higher relative uncertainty than *GEE* and R_{eco}.

Different ecosystem models and optimization methods have obvious effect on R_{eco}, *NEE* and *GEE* estimation and their uncertainties. With a given optimization method, there is not significant distinction between different models. However, with the same model, the uncertainties of any flux components (R_{eco}, *NEE* and *GEE*) obtained by MLE are lower than those obtained by OLS. Uncertainty of R_{eco} calculated from MLE is lower than that from OLS with 15.05 g $C\cdot m^{-2}\cdot a^{-1}$, and the relative uncertainty decreases from 6.58% to 6.13%. Uncertainty of *NEE* calculated from MLE is lower by 9.52 g $C\cdot m^{-2}\cdot a^{-1}$, with relative uncertainty decreasing from 18.7% to 8.97%. *GEE* uncertainty calculated from MLE is also lower than that from OLS by 4.99 g $C\cdot m^{-2}\cdot a^{-1}$, with the relative uncertainty decreasing from 3.24% to 2.39%, which is consistent with the principle of the MLE method, i.e., decreasing the influence of abnormal measurement.

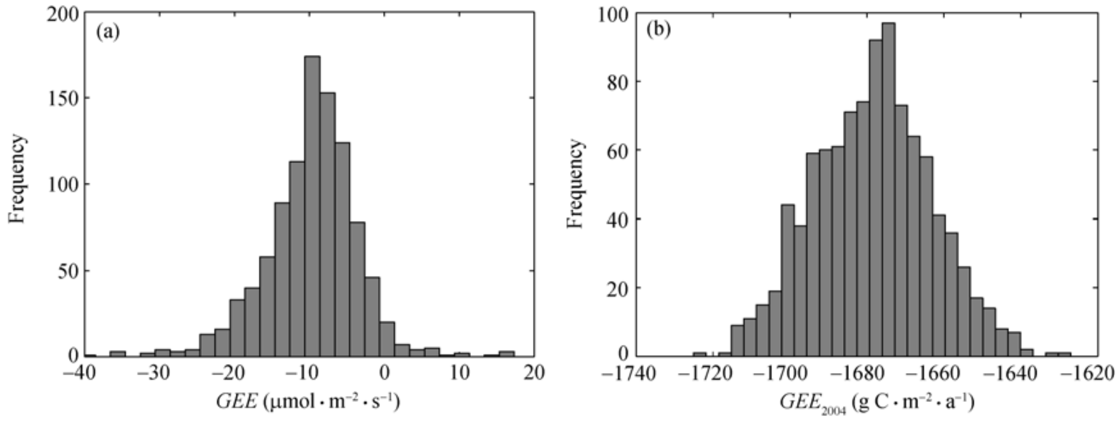

Figure 4 Distribution of *GEE* estimates at different timescales. (a) 30 min timescale (14:30 on 13 July, 2004); (b) annual timescale (2004).

3.4 Uncertainty at different timescales

Uncertainty of CO_2 flux component at different timescales is also calculated (Figure 5). There is a similar tendency for relative uncertainty of different flux components, and uncertainty decreased with the increase of timescale. The relative uncertainties of R_{eco}, *NEE* and *GEE* at annual timescale are 4%—8%, 7%—22%, and 2%—4%, respectively, and 13%—18%, 70%—115%, and 8%—13% at 30 min timescale. The decreases in uncertainty with timescale result from the increase of sample sizes and from the fact that many outliers are smoothed. There is a similar tendency of uncertainty with timescale among R_{eco}, *NEE* and *GEE* ,but *NEE* has an obviously higher uncertainty than the uncertainties of R_{eco} and *GEE*. We may consider that most of *NEE* is valid measurement, which has a greater variation than modeled R_{eco} and *GEE*. Meanwhile, the relative uncertainty from the TW_model is higher than that from the T_model. OLS also leads to a higher relative uncertainty than MLE by 2%. These results are in agreement with the results of parameter and flux components.

4 Conclusions

We present an uncertainty analysis of ecological process parameters and CO_2 flux components (R_{eco}, *NEE* and *GEE*) derived from 3 years' continuous eddy covariance measurements of CO_2 fluxes in subtropical evergreen coniferous plantation, Qianyanzhou in ChinaFlux. We use bootstrapping to quantify the statistical uncertainties of flux components estimations and the daily-differencing approach to analyze the measurement error of CO_2 fluxes measurements. In addition, the effects of different types of ecological process models and parameters optimization methods on key parameters of models and CO_2 flux components are also compared in this paper.

(1) The flux measurement uncertainty and model fit residuals appear to follow a double-exponential distribution, with heavier tails and more prominent central peak, rather than a normal distribution. The non-normal distribution and non-constant variance violate the assumptions that the error is Gaussian and homoscedatic for OLS optimization. For this reason, we argue that it is necessary to implement a new optimization method (such as maximum likelihood estimation) to obtain a reasonable analysis with flux data.

(2) After using OLS and MLE to fit the parameters of two ecosystem process models (T_model, TW_model), we find that the parameters ($R_{r,ref}$, Q_{10}, and P_{max}) estimations obtained by MLE is lower than those obtained by OLS.

(3) The differences between simulated annual R_{eco}, *NEE* and *GEE* derived from MLE and those derived from OLS are 176.5, 79.2, and 92.0 g $C \cdot m^{-2} \cdot a^{-1}$. However, with a given parameter optimization method, the differences between the simulated annual R_{eco}, *NEE* and *GEE* obtained by TW_model and those obteained by the T_model are only 17.8, 5.7, and 4.3 g $C \cdot m^{-2} \cdot a^{-1}$. It suggests that the different types of models may be less important than the optimization methods. Therefore, it is very important to select a suitable error distribution (optimization method) for estimating and evaluating CO_2 flux components.

(4) Bootstrapping algorithm provides the uncertainty analysis on flux components, and the result shows that the model parameters and flux uncertainty appear to follow a normal distribution. With a given model, MLE makes a lower uncertainty (relative uncertainty and uncertainty magnitude) for estimation than OLS. With a given optimization method, the TW_model makes a lower uncertainty for estimation than the T_model. Estimation uncertainty of flux components varied with timescales, that is, the longer the timescale, the lower the uncertainty. The relative uncertainties of annual R_{eco},

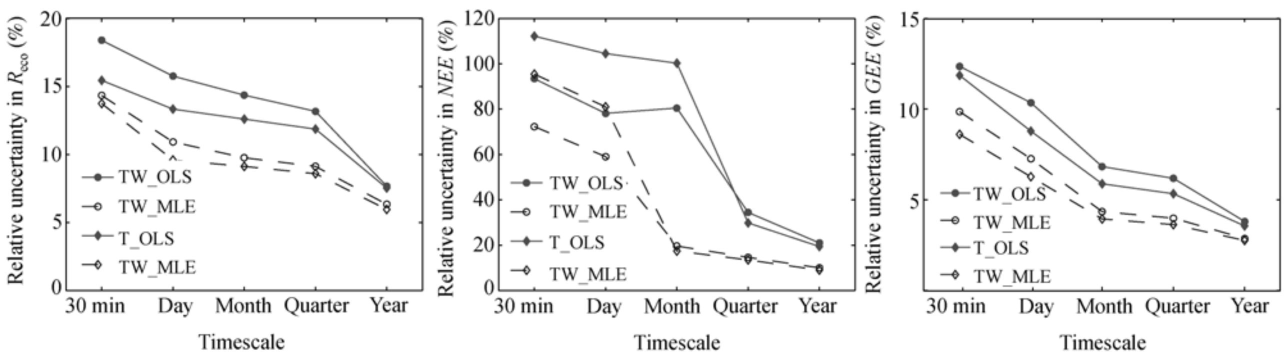

Figure 5 Relative uncertainty estimations of flux components at different timescales.

NEE and *GEE* are 4%—8%, 7%—22%, and 2%—4%, respectively.

(5) In our study, we just analyzed uncertainty of measured *NEE* at one station in Qianyanzhou, and the spatial representative might be limited. Meanwhile, owing to the complexity of eddy covariance technology and the flux measurement, physical mechanism of the distribution of the measurement error will be further studied. The study directions in future are studying the physical mechanism and making further uncertainty analysis with more different models, optimization methods and data sets. How to adopt effective method to reduce the uncertainty of measurement and prediction and really reflect ecosystem-atmosphere carbon exchange condition is also important. Therefore, analysis and studies on long-term observation data in multi-sites are needed to resolve these issues.

The authors thank Qianyanzhou Station, Chinese Ecosystem Research Network (CERN) for kindly providing data sets.

1 Yu G R, Niu D, Wang Q F. Focal issues in the negotiation of United Nations framework convention on climate change (in Chinese). Resour Sci, 2001, 23(6): 10—16

2 Dixon D K, Brown S, Houghton R A, et al. Carbon pools and flux of global forest ecosystem. Science, 1994, 263: 185—190

3 Javis A J, Stauch V J, Schulz K, et al. The seasonal temperature dependency of photosynthesis and respiration in two deciduous forests. Glob Change Biol, 2004, 10(6): 939—950

4 Gove J H, Hollinger D Y. Application of a dual unscented Kalman filter for simultaneous state and parameter estimation in problems of surface-atmosphere exchange. J Geophys Res, 2006, 111: D08S07, doi: 10.1029/2005JD006021

5 Raupach M R, Rayner P J, Barrett D J, et al. Model-data synthesis in terrestrial carbon observation: Methods, data requirements and data uncertainty specifications. Glob Change Biol, 2005, 11: 378—397

6 Hollinger D Y, Richardson A D. Uncertainty in eddy covariance measurements and its application to physiological models. Tree Physiol, 2005, 25: 873—885

7 Richardson A D, Hollinger D Y. Statistical modeling of ecosystem respiration using eddy covariance data: Maximum likelihood parameter estimation, and Monte Carlo simulation of model and parameter uncertainty, applied to three simple models. Agric For Meteorol, 2005, 131: 191—208

8 Richardson A D, Hollinger D Y, George G, et al. A multi-site analysis of random error in tower-based measurements of carbon and energy fluxes. Agric For Meteorol, 2006, 136: 1—18

9 Papale D, Reichstein M, Canfora E, et al. Towards a more harmonized processing of eddy covariance CO_2 fluxes: Algorithms and uncertainty estimation. Biogeosci Discuss, 2006, 3: 961—992

10 Falge E, Baldocchi D, Olson R J, et al. Gap filling strategies for defensible annual sums of net ecosystem exchange. Agric For Meteorol, 2001, 107: 43—69

11 Hui D F, Wang S Q, Su B, et al. Gap-filling missing data in eddy covariance measurements using multiple imputations (MI) for annual estimations. Agric For Meteorol, 2004, 121: 93—111

12 Hagen S C, Braswell B H, Linder E, et al. Statistical uncertainty of eddy flux-based estimates of gross ecosystem carbon exchange at Howland Forest, Maine. J Geophys Res, 2006, 111: D08S03, doi: 10.1029 /2005JD006154

13 Liu Y F, Song X, Yu G R, et al. Seasonal variation of CO_2 flux and its environmental factors in evergreen coniferous plantation. Sci China Ser D-Earth Sci, 2005, 34(Suppl. I): 123—132

14 Liu Y F, Yu G R, Wen X F, et al. Seasonal dynamics of CO_2 fluxes from subtropical plantation coniferous ecosystem. Sci China Ser D-Earth Sci, 2006(Suppl. II): 99—109

15 Li C, He H L, Liu M, et al. The design and application of CO_2 flux data processing system at ChinaFLUX (in Chinese). Geo-Information Sci, 2008, 10(5): 557—565

16 Yu G R, Su X M. Principles of Flux Measurement in Terrestrial Ecosystem (in Chinese). Beijing: High Education Press, 2006. 261—264

17 Lloyd J, Taylor J A. On the temperature dependence of soil respiration. Funct Ecol, 1994, 8: 315—323

18 Fang C, Moncrieff J B. The dependence of soil efflux on temperature. Soil Biol Biochem, 2001, 33: 155—165

19 Davidson E A, Verchot L V, Cattnio J H, et al. Effects of soil water content on soil respiration in forests and cattle pastures of eastern Amazonia. Biogeochemistry, 2000, 48: 53—69

20 Michaelis L, Menten M L. Die Kinetik der Invertinwirkung. Biochem Z, 1913, 49: 333

21 Zhang X Q, Xu D Y. Physiological Model of Forest Growth and Yield (in Chinese). Beijing: Chinese Science and Technique Press, 2002

22 Stock J H, Watson M W. Introduction of Econometrics. New York: Cambridge University Press, 2002

23 Press W H, Teukolsky S A, Vetterling W T, et al. Numerical Recipes in Fortran 77: The Art of Scientific Computing. New York: Cambridge University Press, 1993. 652—655

24 Sitter R V. Bootstrap methods for survey data. Can J Statist, 1992, 20: 135—154

25 Efron B, Tibshirani R J. An Introduction to the Bootstrap. NewYork: Chapman & Hall, 1993

26 Zhang L M. Ecophysiological Controls on Seasonal Variations of Ecosystem Carbon Exchange of Typical Forest Ecosystem along NSTED (in Chinese). Dissertation for Doctoral Degree. Beijing: Graduate University of Chinese Academy of Sciences, 2006. 121

27 Morgenstern K, Black T A, Humphreys E R, et al. Sensitivity and uncertainty of the carbon balance of a Pacific Northwest Douglas-fir forest during an E1Nino/La Nina cycle. Agric For Meteorol, 2004, 123: 201—219

28 Griffis T J, Black T A, Morgenstern K, et al. Ecophysiological controls on the carbon balances of three southern boreal forests. Agric For Meteorol, 2003, 117: 53—71

29 Baldocchi D, Falge E, Gu L, et al. FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. Bull Amer Meteorol Soc, 2001, 82: 2415—2434

30 Lee X H, Fuentes J D, Staebler R M, et al. Long-term observation of the atmospheric exchange of CO_2 with a temperate deciduous forest in southern Ontario. Can J Geophys Res Atmos, 1999, 104: 15975—15984

31 Goulden M L, Munger J W, Fan S M, et al. Measurements of carbon sequestration by long-term eddy covariance: Methods and critical evaluation of accuracy. Glob Change Bio,1996, 2: 169—182