AGRICULTURAL AND FOREST METEOROLOGY XXX (2007) XXX–XXX

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/agrformet

ELSEVIER

Agricultural and Forest Meteorology

Cross-site evaluation of eddy covariance GPP and RE decomposition techniques

Ankur R. Desai[a,1,*], *Andrew D. Richardson*[b], *Antje M. Moffat*[c], *Jens Kattge*[c], *David Y. Hollinger*[d], *Alan Barr*[e], *Eva Falge*[f], *Asko Noormets*[g], *Dario Papale*[h], *Markus Reichstein*[c], *Vanessa J. Stauch*[i]

[a] *Department of Atmospheric and Oceanic Sciences, University of Wisconsin, 1225 W Dayton Street, Madison, WI 53706, USA*
[b] *Complex Systems Research Center, Institute for the Study of Earth, Oceans and Space, University of New Hampshire, Durham, NH, USA*
[c] *Max Planck Institute for Biogeochemistry, Jena, Germany*
[d] *Northern Research Station, U.S. Forest Service, Durham, NH, USA*
[e] *Climate Research Division, Environment Canada, Saskatoon, SK, Canada*
[f] *Max Planck Institute for Chemistry, Mainz, Germany*
[g] *Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, NC, USA*
[h] *University of Tuscia, Viterbo, Italy*
[i] *Federal Office of Meteorology and Climatology (MeteoSwiss), Zurich, Switzerland*

ARTICLE INFO

Article history:
Received 27 July 2007
Received in revised form
14 November 2007
Accepted 23 November 2007

Keywords:
Eddy correlation
Carbon balance
Net ecosystem exchange
GPP
RE

ABSTRACT

Eddy covariance flux towers measure net exchange of land–atmosphere flux. For the flux of carbon dioxide, this net ecosystem exchange (NEE) is governed by two processes, gross primary production (GPP) and a sum of autotrophic and heterotrophic respiration components known as ecosystem respiration (RE). A number of statistical flux-partitioning methods, often developed to fill missing NEE data, can also be used to estimate GPP and RE from NEE time series. Here we present results of the first comprehensive, multi-site comparison of these partitioning methods. An initial test was performed with a subset of methods in retrieving GPP and RE from NEE generated by an ecosystem model, which was also degraded with realistic noise. All methods produced GPP and RE estimates that were highly correlated with the synthetic data at the daily and annual timescales, but most were biased low, including a parameter inversion of the original model. We then applied 23 different methods to 10 site years of temperate forest flux data, including 10 different artificial gap scenarios (10% removal of observations), in order to investigate the effects of partitioning method choice, data gaps, and intersite variability on estimated GPP and RE. Most methods differed by less than 10% in estimates of both GPP and RE. Gaps added an additional 6–7% variability, but did not result in additional bias. ANOVA showed that most methods were consistent in identifying differences in GPP and RE across sites, leading to increased confidence in previously published multi-site comparisons and syntheses. Several methods produced outliers at some sites, and some methods were systematically biased against the ensemble

* *Corresponding author*. Tel.: +1 608 265 9201; fax: +1 608 262 0166.
E-mail address: desai@aos.wisc.edu (A.R. Desai).
[1] Formerly at: National Center for Atmospheric Research, Boulder, CO, USA.
0168-1923/$ – see front matter © 2007 Published by Elsevier B.V.
doi:10.1016/j.agrformet.2007.11.012

UNCORRECTED PROOF

mean. Larger model spread was found for Mediterranean sites compared to temperate or boreal sites. For both real and synthetic data, high variability was found in modeling of the diurnal RE cycle, suggesting that additional study of diurnal RE mechanisms could help to improve partitioning algorithms.

© 2007 Published by Elsevier B.V.

1. Introduction

The terrestrial component of the global carbon cycle can be divided in two large and opposing terms, both of which represent aggregated ecosystem processes: gross primary production (GPP) and total ecosystem respiration (RE). The order of magnitude smaller imbalance between these two fluxes, termed net ecosystem exchange (NEE), is considered to be the primary source of observed interannual variability in atmospheric accumulation of carbon dioxide (CO_2) (Peylin et al., 2005). Furthermore, understanding how plant and soil processes impact this interannual variability requires quantifying GPP and RE. However, it is currently not possible to obtain direct, integrated observations of either GPP or RE, because these processes represent a multitude of responses by a combination of autotrophic and heterotrophic organisms. Scaling from chamber level measurements to canopy level is labor intensive and fraught with high sampling uncertainty.

The eddy covariance (EC) technique is the well-established method to directly measure flux and NEE over a fetch larger than typical plot level measurements (Baldocchi, 2003). Gaps in NEE time series are inevitable due to operational and micrometeorological constraints. Numerous methods have been developed to fill the gaps due to observational and micrometeorological constraints, and many of these also decompose NEE into GPP and RE (Falge et al., 2001). In most of the methods, errors in estimation of RE offset errors in GPP, so gap filling of NEE by modeling GPP and RE has been largely successful (Moffat et al., ~~in press~~).

Methods to partition NEE to its component fluxes, GPP and RE, have also been developed independent of gap-filling techniques as a way to assess carbon pathways in ecosystems. At present, there is no standard method commonly in use (Reichstein et al., 2005; Stoy et al., 2006). While many partitioning methods typically rely on the concept of zero GPP at night and strong correlation of GPP and RE to environmental driving variables, such as temperature, water availability and solar radiation (Law et al., 2002), newer techniques, such as neural networks, which have few underlying assumptions regarding these relationships, have been developed and are evaluated here. We also investigated process-based ecosystem model inversion and advanced data assimilation techniques which have only recently been developed.

Despite advances in NEE partitioning, direct evaluation of GPP and RE estimates has been scant. Previous studies have tested multiple methods at a few sites (Stoy et al., 2006) or a few methods at many sites (Falge et al., 2001; Law et al., 2002; Richardson et al., 2006a; Reichstein et al., 2005). Analyzing NEE time series from a boreal transition forest, Hagen et al. (2006) reported that GPP estimates for a given year could vary by over 100 g C m^{-2} depending on the partitioning algorithm (neural network vs. physiologically based) and fitting method (maximum likelihood vs. ordinary least squares) used. Evaluation of GPP and RE at multiple sites with multiple methods has not been performed. There is great interest in performing cross-site comparison of GPP and RE. Without an evaluation of GPP and RE methods across a range of sites, investigator-reported values of GPP and RE for individual sites cannot be reasonably used to compare values across multiple sites because it is not known how the partitioning method employed may affect the result.

The goal of this article is not to discuss mechanistic evaluation of GPP and RE. To do this requires independent flux observations from chambers, biometry, and models or inversions, each of which is subject to its own set of errors and uncertainties. Instead, our focus is on assessing the role of model selection and data gaps on variability in GPP and RE estimates derived from NEE time series. To accomplish this assessment, we evaluated 23 different partitioning methods, using 10 site years of CO_2 flux data. These data, originally compiled for a gap-filling intercomparison (Moffat et al., ~~in press~~), come primarily from temperate forests sites in Europe. Though not all kinds of ecosystems are tested, the sites chosen span a reasonable range of variability seen in flux tower time series.

Questions motivating this research are

1. What is the inherent variability in estimated GPP and RE for any single site as a function of method, and what does this imply for giving uncertainty bounds on GPP and RE values from any one method?
2. Is within site variability of derived GPP and RE as a function of partitioning method smaller than typical interannual variability in GPP and RE (~10% of 100 gC m^{-2} $year^{-1}$, Richardson et al., 2007)?
3. Are some methods more sensitive to data gaps than others in terms of mean variability? Do gaps induce any systematic biases?
4. Does choice of partitioning method alter understanding of differences in seasonal and diurnal variability of GPP and RE, or cross-site rankings of annual sums of these component fluxes? Are certain methods systematically biased across the sites with respect to the ensemble mean of GPP or RE?

Though independent evaluation of GPP and RE is not performed here, a preliminary test of method fidelity can be done by testing against synthetic data (Stauch and Jarvis, 2006). Prior to comparison of methods against observed data, we investigated whether methods could accurately estimate GPP and RE from NEE generated by a reasonably complex, complete and well-tested ecosystem model, BETHY (Knorr and

UNCORRECTED PROOF

Please cite this article in press as: Desai, A.R. et al., Cross-site evaluation of eddy covariance GPP and RE decomposition techniques, Agric. Forest Meteorol. (2007), doi:10.1016/j.agrformet.2007.11.012

Kattge, 2005). To further simulate observation conditions, artificial noise mimicking the random noise statistics of EC observed NEE (Richardson et al., 2006b) was added to this synthetic NEE. While this is not a perfect test, it did allow for evaluation of partitioning methods performance relative to known ''truth'', which, as noted above, is not possible with current field measurement technology. Further, by adding artificial gaps to the synthetic data, we evaluated method bias induced by gaps.

2. Methods

2.1. Flux partitioning methods

GPP and RE estimates from a total of 23 different methods participated (Table 1). These approaches are described fully by Moffat et al. ~~(in press)~~ and the citations noted in Table 1, but a brief overview is given here.

The largest batch of partitioning methods was of the non-linear regression methods. These methods rely on correlating nighttime NEE, representing RE, to temperature, time and moisture variables, and daytime NEE, representing the combination of GPP and RE, to temperature and radiation variables. The primary differences among methods are choice of functional form, meteorological forcing variables, fixed vs. free parameters, parameter time dependence, time window size, statistical goodness-of-fit test, and whether regression is done first on nighttime, daytime, or all NEE. These details are found in Moffat et al. ~~(in press)~~.

Lookup table and diurnal course type methods formed the second largest batch of partitioning methods. Lookup tables rely on binning NEE data by one or more of the forcing variables across a number of time periods (Falge et al., 2001). Extrapolation with nighttime data against air temperature and soil temperature or daytime data with a light intercept (use daytime flux and extrapolate to zero incoming PAR) is used to compute RE while GPP is solved as a residual. Diurnal course methods perform multiple-day ensemble averaging across suitable time windows.

A number of alternative statistical techniques were also tested on the datasets. B365 is based on BETHY, a soil-vegetation-atmosphere-transport (SVAT) type ecosystem model (Knorr and Kattge, 2005). The model is forced with the observed meteorology. The Markov Chain Monte Carlo (MCMC) technique, a Bayesian parameter estimation algorithm, is applied against the NEE data to optimize model parameters (Knorr and Kattge, 2005).

The SPM technique estimates a three dimensional hypersurface from the observations to describe the net CO_2 exchange as a continuous function of radiation, temperature and time (Stauch and Jarvis, 2006). As such, it can be viewed as

Table 1 – List of methods used to derive GPP and RE for all sites. Detailed descriptions can be found in Moffat et al. ~~(in press)~~ or the noted citation. Abbreviations used by Moffat et al. ~~(in press)~~ are noted in italics

Abbreviation	Description	Citation
Non-linear regression		
NA (*NLR_AM*)	Noormets model	Noormets et al. ~~(in press)~~
NE (*NLR_EM*)	Eyring respiration model	Desai et al. (2005)
NFA (*NLR_FM_AD*)	Absolute deviation model	Richardson et al. (2006a)
NFO (*NLR_FM_OLS*)	Ordinary least squares model	Richardson et al. (2006a)
NFW[a]	Weighted absolute deviation model	Richardson et al. (2006a)
NLI	Light intercept based regression	Falge et al. (2001)
NLT (*NLR_LM*)	Air temperature based regression	Falge et al. (2001)
NLS	Soil temperature based regression	Falge et al. (2001)
NC1 (*NLR_FCRN*)	Multi timescale regression	Barr et al. (2004)
NC2[b]	Multi timescale regression	Barr et al. (2004)
MR1	Long term air temperature regression	Reichstein et al. (2005)
MR1R	Robust long term air temperature	Reichstein et al. (2005)
MR2	Short term air temperature regression	Reichstein et al. (2005)
MR2R	Robust short-term air temperature	Reichstein et al. (2005)
Lookup tables/mean diurnal course		
NLID	Diurnal course with light intercept	Falge et al. (2001)
NLIL	Lookup table with light intercept	Falge et al. (2001)
NLTD (*MDV*)	Diurnal course with air temperature	Falge et al. (2001)
NLTL (*LUT*)	Lookup table with air temperature	Falge et al. (2001)
NLSD	Diurnal course with soil temperature	Falge et al. (2001)
NLSL	Lookup table with soil temperature	Falge et al. (2001)
Other methods		
B365 (*BETHY_ALL*)	Ecosystem model inversion	Knorr and Kattge (2005)
SPM (*SPM*)	Semi-parametric method	Stauch and Jarvis (2006)
UKF[b] (*UKF_LM*)	Unscented Kalman filter	Gove and Hollinger (2006)
ANN (*ANN_PS*)	Artificial neural network	Papale and Valentini (2003)
ANNS[a]	Artificial neural network with soil moisture	Papale and Valentini (2003)

[a] Method used only for synthetic analysis.
[b] Method not used in synthetic analysis.

Please cite this article in press as: Desai, A.R. et al., Cross-site evaluation of eddy covariance GPP and RE decomposition techniques, Agric. Forest Meteorol. (2007), doi:10.1016/j.agrformet.2007.11.012

both a non-linear regression without a prescribed functional form or a lookup table without binning the data. The underlying semi-parametric (multidimensional) relationships are described by cubic Hermite splines. The estimation of the respiration component is based on the light independent response of the hypersurface, i.e., the SPM partitioning scheme makes use of all NEE data. The gross CO_2 uptake is then calculated as the difference between the estimates NEE and RE (Stauch, 2007).

UKF is a dual unscented Kalman filter recursive predictor-corrector method used to adjust the parameters of non-linear equations (Gove and Hollinger, 2006). NEE and other observed state variables, that are inherently noisy, are used to update predictions of the state by a non-linear process model. Continuous time series of optimal model state, model parameters and uncertainty are provided. In the dual scheme, two filters are run in parallel for state and parameter estimation, respectively.

ANN is an artificial neural network based method (Papale and Valentini, 2003). ANN is essentially a non-linear regression that mimics neural learning patterns and relies on the data to discover the inherent functional relationships between driver data and NEE (Moffat et al., ~~in press~~). Additionally, ANN_S was used in the synthetic data analysis to test the role of soil moisture as an additional predictor variable.

2.2. *Synthetic model–model comparison*

An initial comparison of the GPP and RE flux partitioning methods was performed by evaluating their ability to retrieve GPP and RE from synthetic data produced by an ecosystem carbon cycle model. We used the BETHY model (Knorr and Kattge, 2005) to simulate GPP and RE (with NEE then equal to the residual) of a typical mid-latitude European forest forced with observed meteorology, using model parameter values appropriate for the site in question (DE3_2000, a mixed forest). Methods did not know which particular site was being simulated. To further mimic real-world conditions, noise typical of real NEE measurements (Hollinger and Richardson, 2005; Richardson et al., 2006b) was added to the synthetic NEE data, which, along with meteorological drivers (air temperature, soil temperature, PAR and soil moisture) was provided to participants. The added noise was randomly drawn from a double exponential distribution whose magnitude was proportional to the measured flux as described in Hollinger and Richardson (2005).

A subset of method investigators tested their models on the synthetic NEE data. Two other methods, NFW and ANN_S were tested with synthetic data but not with real data. These methods were used to test an alternate error model for the NF* series of methods and adding soil moisture to the neural network, respectively. Output GPP and RE from the methods were then compared to the original BETHY model GPP and RE using a variety of statistical tests. This test did not reveal which is the best method for deriving GPP and RE, but rather provided a simple test of variability of derived GPP and RE against a known modeled value with noise.

2.3. *Observed data analysis*

After the model–model analysis, model–data analysis was performed using observed flux data. Flux tower NEE data from six sites (10 site years) were taken from the CarboeuropeIP database (Table 2). These datasets were the same as those used in the NEE gap-filling comparison project (Moffat et al., ~~in press~~). The sites spanned a range of European forests and climates, from Mediterranean to boreal. Meteorological forcing data of air temperature, soil temperature and incident photosynthetic active radiation (PAR) for each site were gap filled using a variety of interpolation techniques as described by Moffat et al. ~~(in press)~~. All NEE data were screened and filtered with a standardized method (Papale et al., 2006), leading to 70–90% data availability in daytime and 30–40% at night.

Methods derived GPP and RE were compared against one another for each site at the annual, monthly and diurnal timescales. Deviation from mean plots in absolute and relative values was computed to look for model-based variability in GPP and RE. Median, interquartile range (IQR) and max–min statistics were the primary assessment techniques to look for ensemble, typical model, outlier model performance statistics. Ranked statistics and ANOVA analysis on method by site were performed to test for ranked coherence of sites as a function of method and for systematic biases in methods as a function of site.

2.4. *Artificial gap scenarios*

To further test method robustness under real observation condition, artificial data gaps were added to the NEE data. A total of 10 scenarios were used based on the mixed gap set described in Moffat et al. ~~(in press).~~ Using a combination of gaps of varying lengths (from individual half hours to a single 12-day period), roughly 10% of the real NEE measurements were removed from each time series. Both the real data and the synthetic data were subject to these gap scenarios and the methods produced new GPP and RE estimates for each site year/gap scenario combination, which were then compared to the original (no artificial gap) derived GPP and RE.

UNCORRECTED PROOF

Table 2 – Site names, major species, years of analysis and locations used in this analysis

Site	Location	Species	Years	Lat (°N)	Lon (°E)	Reference
be1	Viesalm, Belgium	*Fagus sylvatica, Pseudotsuga menziesii*	2000, 2001	50.30	5.98	Aubinet et al. (2001)
de3	Hainich, Germany	*Fagus sylvatica*	2000, 2001	51.07	10.45	Knohl et al. (2003)
fi1	Hyytiala, Finland	*Pinus sylvestris*	2001, 2002	61.83	24.28	Suni et al. (2003)
fr1	Hesse, France	*Fagus sylvatica*	2001, 2002	48.67	7.05	Granier et al. (2000)
fr4	Puechabon, France	*Quercus ilex*	2002	43.73	3.58	Rambal et al. (2004)
it3	Roccarespampani, Italy	*Quercus cerris*	2002	42.40	11.92	Tedeschi et al. (2006)

Please cite this article in press as: Desai, A.R. et al., Cross-site evaluation of eddy covariance GPP and RE decomposition techniques, Agric. Forest Meteorol. (2007), doi:10.1016/j.agrformet.2007.11.012

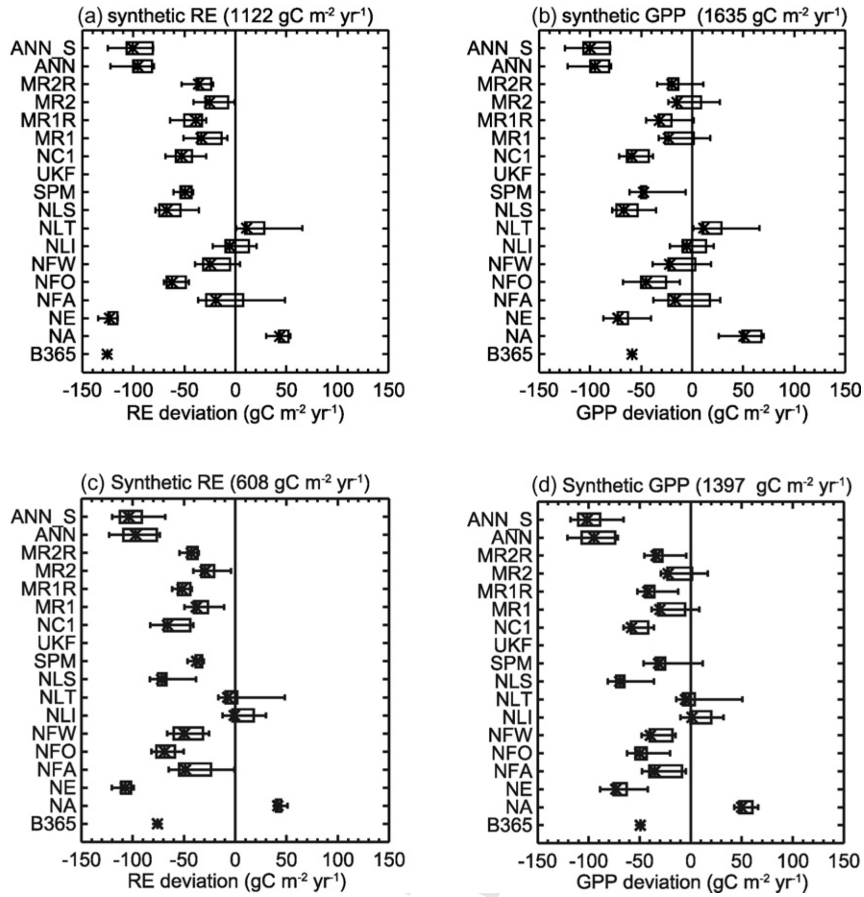

Fig. 1 – Deviation (star) from modeled (a) annual RE, (b) annual GPP, (c) May-Sep RE, and (d) May-Sep GPP for each method that produced GPP and RE from the noisy synthetic NEE dataset produced by the BETHY model. Most methods were biased low against the model RE and GPP. Effect of gaps, shown by interquartile range (box) and total range (line), was to skew GPP and RE slightly positive for most methods, a small effect that has no simple explanation. Method B365 had zero variation as it did not perform a gap sensitivity test.

3. Results

3.1. *Synthetic flux analyses*

The methods generally were able to retrieve BETHY model driven GPP and RE given artificially noisy NEE and gap-filled meteorological forcing to within 100 g C m^{-2} year^{-1} (Fig. 1a and b). In terms of annual RE and GPP, mean deviation was −47 g C m^{-2} year^{-1} (range −126 to +43) for RE and −35 g C m^{-2} year^{-1} (range −100 to +51) for GPP, and all but two methods were biased low against the ‘‘true’’ GPP and RE. The Markov Chain Monte Carlo version of BETHY (B365)) had the largest annual RE bias, while the ANN method had the largest GPP bias. In both cases, the smallest bias was found with NLI. Mean absolute errors were 54 g C m^{-2} year^{-1} (range +5 to +126) for RE and 44 g C m^{-2} year^{-1} (range +5 to +100) for GPP. In relative terms, methods were within 4.8% for RE and 2.7% for GPP. Most of the biases occurred during the summer season (Fig. 1c and d), as might be expected given it is the season when fluxes were largest in absolute magnitude. Wintertime fluxes were generally well modeled by all methods with low bias.

The 10 artificial gap scenarios added additional source of variability to the RE and GPP retrieval, with an IQR average of 19 g C m^{-2} year^{-1} (range +9 to +36) for RE and 21 g C m^{-2} year^{-1} (range +5 to +41) for GPP. For individual methods, max–min variability across the 10 different scenarios averaged 49 g C m^{-2} year^{-1} (range +33 to +66) for RE and 40 g C m^{-2} year^{-1} (range +19 to +85) for GPP. The NFA model had the largest variability with respect to gaps for both GPP and RE. NE and SPM methods had the smallest gap variability for RE IQR and max–min, respectively, while SPM and NC1 were the smallest for GPP IQR and max–min. While most methods were negatively biased with respect to synthetic GPP and RE, adding gaps to NEE tended to increase method GPP and RE, leading to a smaller bias against synthetic RE and GPP for most models, though this is likely a coincidence. This effect is in contrast to the real data scenarios, where gaps just increased variability in a non-systematic fashion.

We looked at the correlation of GPP and RE predicted by BETHY with predictions of each of the partitioning methods at both the hourly and daily timescale. Correlation of method RE to BETHY RE at hourly scales was significantly improved when aggregated to the daily scale (Fig. 2). The analysis of the observed data showed that this is very likely due to choice of RE diurnal cycle representation in the methods. Poor hourly correlation was found with NFA, NFW, NFO and NC1 methods.

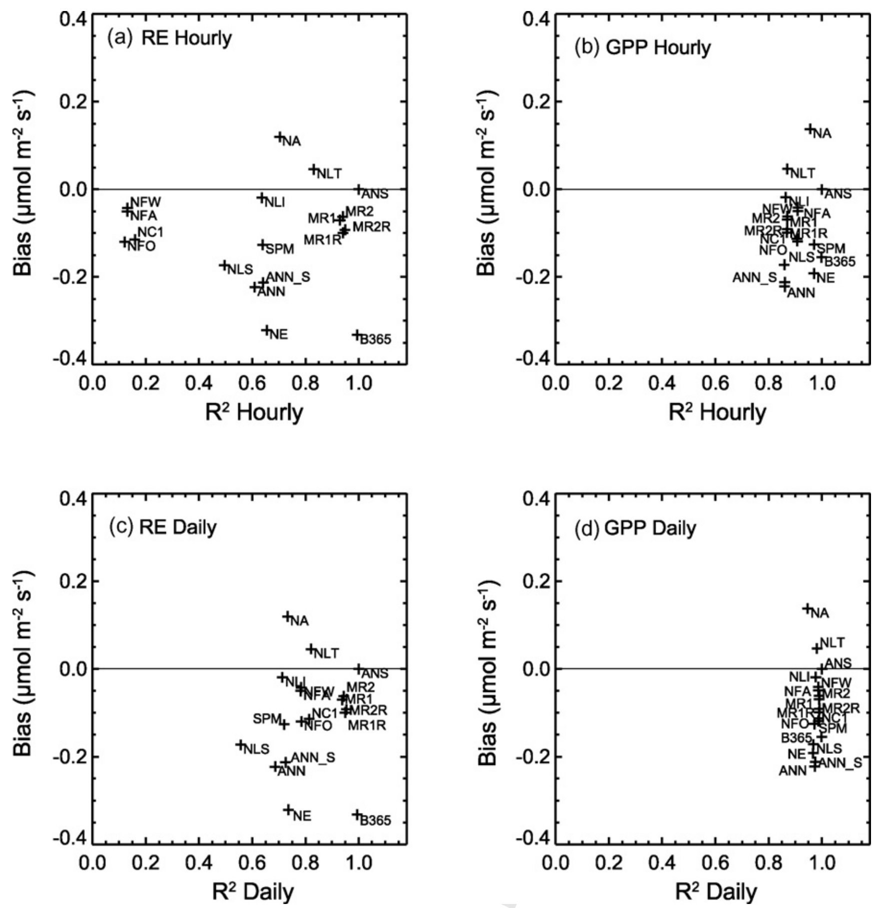

Fig. 2 – Comparison of correlation coefficient (R^2) to mean annual half-hourly bias for (a) RE at hourly scales, (b) GPP at hourly scales, (c) RE at daily scales and (d) GPP and daily scales for each method against the synthetic GPP and RE dataset. Weak correlation for RE at hourly scales disappeared at the daily scale. GPP correlation was strong at all timescales. Parameter inversion of the synthetic model (B365) produced high correlation but a large negative bias.

All methods perform better at the daily scale, some more than others. NLS has the lowest correlation to synthetic RE at the daily scale. For GPP, strong correlation was found for all methods on both the hourly and daily scale.

The MCMC parameter inversion of the BETHY model had the highest correlation to the synthetic data GPP and RE, which could be expected given the 1:1 correspondence in model equations, but not the lowest bias (Fig. 8). NLI had the lowest bias for RE and GPP, but was in the middle for the pack on correlation. The MR1, MR1R, MR2 and MR2R suite of methods had generally strong performance in both bias and correlation.

3.2. *Partitioning method variability*

When the methods were applied to real observed NEE, variability by partitioning method in GPP and RE was found to be relatively small with respect to annual totals (Fig. 3). IQR of GPP and RE from all the methods was typically less than 10% of the annual sum of GPP or RE for any particular site. For RE (Fig. 3a), the IQR averaged 108 g C m^{-2} $year^{-1}$. This translates to a mean variability of 9.8% (5.9–12.3%) of the annual RE. However, outliers across some methods pushed the total mean range (max–min) to 366 g C m^{-2} $year^{-1}$. For GPP (Fig. 3b), very similar ranges are seen in IQR but fewer outliers led to a smaller max–min range. Mean IQR was 104 g C m^{-2} $year^{-1}$ and 7.0% in relative terms, while max–min range averaged 314 gC m^{-2} $year^{-1}$ of annual GPP.

Large outliers for some methods existed for several sites, especially the Mediterranean forests (IT3 and FR4). Sites with the largest spread in IQR or max–min range for both RE and GPP were the Mediterranean sites, FR1_2001 and IT3_2002, with max–min range exceeding 450 g C m^{-2} $year^{-1}$. Deciduous forest FI1_2001 and broadleaf evergreen FR4_2002 had the smallest range across methods, less than 210 g C m^{-2} $year^{-1}$ for max–min in RE and 180 g C m^{-2} $year^{-1}$ max–min for GPP. These numbers could be considered an estimate of the upper bound of uncertainty expected due to model selection.

GPP/RE ratios (Fig. 3c) typically showed smaller variation across methods, with a mean IQR relative variation of 2.5% (range 1.0–4.2%). Max–min ranges were also smaller, with mean max–min of 5.8% (3.7–10.2%). These results are similar to the variability found in gap filling of NEE (Moffat et al., ~~in press),~~ Though large IQR was found for IT3_2002, the site had the smallest range of GPP/RE, reflecting the role of compensating errors in GPP and RE for models that are inverted against a given NEE.

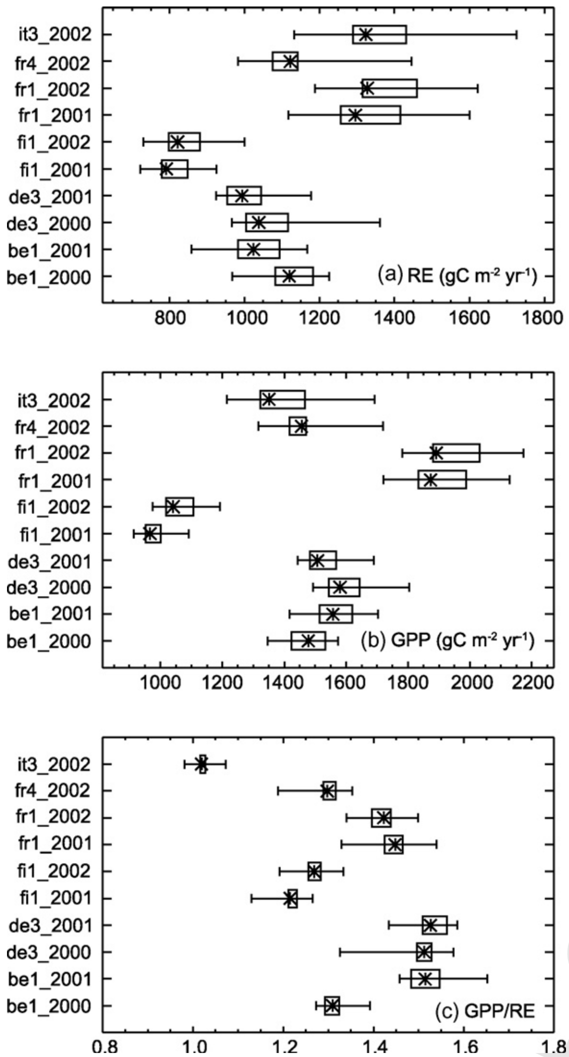

Fig. 3 – Median (star), interquartile range (box) and total range (line) of annual (a) RE, (b) GPP and (c) GPP/RE ratio for each site as a function of GPP/RE method. Uncertainty was greater in RE and also for Mediterranean site GPP and RE. Most methods were within ~100 gC m^{-2} year^{-1} of each other for GPP and RE, around 10% of annual GPP and RE, though large outliers existed at many sites.

3.3. *Biases and cross-site rankings*

Though good agreement was found across model GPP and RE, several methods were found to be biased high or low with respect to the ensemble mean. Though the ensemble mean is not necessarily the ''correct'' or ''true'' GPP or RE, the model deviations provide a way to classify methods into groups and identify any systematic outliers. We conducted an analysis of variance (ANOVA), with 'partitioning method' as a main effect and 'site' as a blocking factor, and then used a Bonferroni multiple comparison test to identify groups of partitioning methods that produced similar results (Tables 3 and 4; black bars indicate groups of methods that were not significantly different from one another in the multiple comparison test). This analysis indicated six different (but largely overlapping) groups of methods for GPP, and seven groups for RE. In both cases, the UKF (which produced the highest estimates of both GPP and RE) was in its own group, and thus significantly different from all other methods. The NE method produced the lowest estimates of both GPP and RE, but was always grouped with a number of other methods, including NFA, NLSLS, and ANN, indicating that these methods did not produce results that were significantly different from each other according to the ANOVA analysis. For RE, groups c and f (for GPP, groups c and e) included 20 of the 23 methods used (all except NE, NA, and UKF).) Biases evident in RE (Table 3) were generally identical to biases in GPP (Table 4), which could be expected given the covariance between GPP and RE the methods produce for a given NEE (i.e., for a given NEE, and RE estimated by a particular method, then by definition GPP = NEE + RE). In general, differences at the annual scale were also reflected at the seasonal scale (data not shown).

In spite of the effects of method biases and variability, cross-site rankings of sites due to partitioning method were surprisingly robust (Tables 5 and 6). Methods were unanimous in selecting sites FI1_2001 and FI1_2002 as the sites with the smallest GPP and RE, and site FR1_2002 with the largest GPP and RE. However, the ANOVA showed that the ''site'' and ''method'' effects are largely additive (i.e., the model residual, which by default includes any ''method'' × ''site'' interaction effect, was small, less than 2% of the total variance), implying that while each method is internally consistent in its ranking of sites of highest and lowest GPP or RE, comparisons of one site with one method to another site with another method is likely to be inaccurate unless the ANOVA results show that the two methods produce statistically similar comparisons (i.e., same letter grouping in Tables 3 and 4). Thus, an important result is that partitioning method must be taken into account when comparing GPP and RE across sites. On the other hand, if all sites had GPP and RE derived from the same method (at least among the ones tested here), the rankings of which sites had highest and lowest GPP or RE should be generally insensitive to which method one chooses.

3.4. *Gap sensitivity*

Sensitivity of methods to data gaps was significantly smaller than sensitivity of method choice for GPP and RE. For each method at each site, GPP and RE were computed with 10 artificial gap scenarios and compared to the GPP and RE computed by the method for data with no artificial gaps. The relative variation on annual GPP and RE due to the 10 gap scenarios ranged from 5 to15% for RE and 4 to 10% for GPP across the various partitioning methods (Fig. 4).

Several methods, in particular UKF, SPM and NLID, were especially sensitive to gaps in that GPP and RE estimates varied widely among the different artificial gap scenarios (Fig. 4). For these methods, gaps tended to reduce GPP and RE by less than 10% compared to the no artificial gap scenario. NLS and NE had the smallest gap sensitivities for RE, while NLS and NC2 were smallest for GPP. The median deviation across all gap scenarios for most methods was at or near zero, implying that the addition of 10% artificial gaps did not generally add a systematic bias to GPP and RE.

UNCORRECTED PROOF

Please cite this article in press as: Desai, A.R. et al., Cross-site evaluation of eddy covariance GPP and RE decomposition techniques, Agric. Forest Meteorol. (2007), doi:10.1016/j.agrformet.2007.11.012

Table 3 – Ranking of method RE for each site and ANOVA correspondence statistics for significant differences across methods for all sites

ANOVA a	b	c	d	e	f	g	Method/Site	be1_2000	be1_2001	de3_2000	de3_2001	fi1_2001	fi1_2002	fr1_2001	fr1_2002	fr4_2002	it3_2002
						■	NE	2	2	2	1	1	1	1	1	9	1
					■	■	NFA	3	3	3	8	6	5	2	2	2	5
				■	■	■	NLSL	7	10	4	3	3	2	6	11	16	2
				■	■	■	NLSD	9	9	8	2	10	8	3	5	13	3
				■	■	■	NLS	8	11	5	6	5	3	8	9	21	4
				■	■	■	ANN	11	4	11	11	4	14	9	3	10	7
			■	■	■	■	NC1	4	6	15	12	12	7	4	8	8	6
		■	■	■	■		NFO	5	7	12	13	13	13	5	4	5	13
		■	■	■	■		SPM	10	5	13	9	20	12	7	12	7	8
		■	■	■	■		NC2	6	8	17	14	11	9	11	7	11	12
		■	■	■	■		B365	1	1	20	18	18	4	13	15	4	20
		■	■	■	■		MR2R	12	12	16	16	9	17	10	6	14	9
		■	■	■			NLIL	14	14	1	10	2	10	20	21	3	18
		■	■	■			NLTL	16	18	6	5	15	11	17	18	17	10
	■	■	■	■			NLTD	21	20	9	4	19	18	16	16	12	11
	■	■	■	■			MR1R	17	13	18	17	16	20	12	10	15	15
	■	■	■	■			NLT	18	19	7	7	17	16	18	17	20	14
	■	■	■	■			MR2	13	16	19	19	14	19	14	13	18	16
	■	■	■	■			NLID	15	15	14	21	7	6	22	19	1	17
	■	■	■				MR1	19	17	21	20	21	21	15	14	19	19
	■	■					NLI	22	21	10	15	8	15	21	22	6	22
	■						NA	20	23	22	22	22	22	19	20	22	21
■							UKF	23	22	23	23	23	23	23	23	23	23

Methods that share a black box were not significantly different from each other in this test. Lower rankings equal lower calculated RE. A handful of sites had a consistent low (NE, NFA) or high (NA, UKF) bias, but most did not.

Table 4 – Ranking of method GPP for each site and ANOVA correspondence statistics for significant differences across methods for all sites

ANOVA a	b	c	d	e	f	Method/Site	be1_2000	be1_2001	de3_2000	de3_2001	fi1_2001	fi1_2002	fr1_2001	fr1_2002	fr4_2002	it3_2002
					■	NE	2	2	6	4	1	1	1	1	8	1
				■	■	NFA	3	3	7	3	9	9	2	2	2	7
			■	■	■	NLSL	6	11	4	5	3	2	6	10	17	2
			■	■	■	NLSD	8	5	8	1	7	5	8	7	13	3
			■	■	■	NLS	4	10	2	6	5	3	7	9	20	4
			■	■	■	NC1	5	7	13	11	12	7	4	3	10	5
			■	■	■	ANN	11	4	11	10	4	13	9	4	9	6
			■	■	■	SPM	10	6	12	12	11	8	3	11	7	8
			■	■	■	NFO	7	9	15	9	14	14	5	5	4	14
		■	■	■	■	NC2	9	8	16	14	13	11	11	8	11	12
		■	■	■		B365	1	1	18	18	17	6	15	16	6	20
		■	■	■		MR2R	12	12	17	15	10	18	10	6	14	10
		■	■	■		NLIL	13	15	1	13	2	10	20	19	3	18
		■	■	■		NLTL	16	19	5	7	16	12	16	18	18	9
		■	■			NLTD	19	16	10	2	20	17	17	15	12	11
		■	■			NLT	17	18	3	8	19	16	18	17	21	13
	■	■	■			NLID	14	13	14	20	6	4	22	21	1	17
	■	■	■			MR1R	20	14	19	17	18	20	12	12	15	15
	■	■	■			MR2	15	17	20	19	15	19	13	13	16	16
	■	■				NLI	22	21	9	16	8	15	21	22	5	22
	■	■				MR1	21	20	21	21	21	22	14	14	19	19
	■					NA	18	23	22	22	22	21	19	20	22	21
■						UKF	23	22	23	23	23	23	23	23	23	23

Methods that share a black box were not significantly different from each other in this test. Lower rankings equal lower calculated GPP. A handful of sites had a consistent low (NE, NFA) or high (NA, UKF) bias, but most did not.

Please cite this article in press as: Desai, A.R. et al., Cross-site evaluation of eddy covariance GPP and RE decomposition techniques, Agric. Forest Meteorol. (2007), doi:10.1016/j.agrformet.2007.11.012

Table 5 – Ranking of site RE by each method and ANOVA statistics showing significant differences across sites as classified by all methods

Method/Site	fi1_2001	fi1_2002	de3_2001	be1_2001	de3_2000	fr4_2002	be1_2000	fr1_2001	it3_2002	fr1_2002
a								■	■	■
b						■	■			
c				■	■					
d			■	■						
e	■	■								
B365	2	1	5	3	7	6	4	8	10	9
NA	1	2	3	4	6	7	5	8	9	10
NE	1	2	4	3	5	7	6	8	9	10
NFA	1	2	4	3	5	6	7	8	10	9
NFO	1	2	4	3	5	6	7	8	10	9
NLID	1	2	5	6	4	3	7	10	8	9
NLIL	1	2	4	6	3	5	7	8	9	10
NLI	1	2	3	6	4	5	7	8	9	10
NLTD	1	2	3	5	4	6	7	9	8	10
NLTL	1	2	3	5	4	6	7	9	8	10
NLT	1	2	3	5	4	6	7	9	8	10
NLSD	1	2	3	4	5	7	6	8	9	10
NLSL	1	2	3	5	4	7	6	9	8	10
NLS	1	2	3	5	4	7	6	8	9	10
SPM	2	1	4	3	5	6	7	8	9	10
UKF	1	2	4	3	6	7	5	8	10	9
NC1	1	2	4	3	5	7	6	8	9	10
NC2	1	2	4	3	5	7	6	8	10	9
MR1	1	2	3	4	5	6	7	8	10	9
MR1R	1	2	3	4	5	6	7	8	10	9
MR2	1	2	3	4	5	6	7	8	10	9
MR2R	1	2	3	4	5	7	6	8	10	9
ANN	1	2	4	3	5	6	7	8	9	10
Max	2	2	5	6	7	7	7	10	10	10
Min	1	1	3	3	3	3	4	8	8	9

Sites that do not share a black box had significantly different RE according to the partitioning methods. This analysis indicates that robust comparison across sites is possible given the strong correspondence in site rankings. Largest disagreements were found for sites de3_2000 and fr4_2002.

3.5. *Seasonal and diurnal trends*

Seasonal and diurnal analyses of GPP and RE are typically used for analysis of environmental controls on photosynthesis and respiration. Ideally, this kind of analysis would not be affected by the choice of NEE partitioning method. However, given the differences among the partitioning methods at the annual timescale, we expected the methods to differ in their seasonal and diurnal patterns of NEE partitioning.

For seasonal analysis, the differences among methods were found to be generally small among 10 site years analyzed; i.e., all methods yielded relatively consistent estimates of the seasonal pattern (Figs. 5 and 6). Here, to increase visual clarity, only one year for each of the six unique sites is shown. For RE (Fig. 5), methods generally had strong agreement on the course of monthly RE, though this was more true for the non-Mediterranean sites. Methods were consistent in showing decreased RE in July for BE1_2000 and peak respiration in May for DE3_2002 (though with greater variability given the large outlier for August). Though the large decrease in RE in July-August for FR4_2002 was replicated by most partitioning methods, there was large uncertainty in its magnitude across all the methods. Results for monthly GPP have similar results with fewer outliers (Fig. 6). Methods portrayed what appear to be typical evergreen and deciduous trends in GPP (Falge et al., 2002; Law et al., 2002). Greater variation among methods was seen again in the Mediterranean sites, FR4_2002 and IT3_2002, perhaps indicating less of a consensus on the environmental controls over seasonal patterns of variation in these ecosystems compared to temperate or boreal systems. Additionally, large gaps are found in IT3_2002. Finally, much of the variability in outliers is due to one or two methods, most notably UKF. Methods NA, NLID, and B365 also tended to be positively biased from the ensemble mean.

Summer diurnal patterns for RE had far less coherence across methods (Fig. 7). This lack of agreement stemmed from both (1) choice of air temperature vs. soil temperature as primary control of respiration (the latter dampening high frequency variability) and (2) high frequency filters for RE present in some of the methods. Methods B365, NA, NLTD, NLTL, NLTR, UKF, MR1 and MR1R had more pronounced diurnal courses for RE than the other methods. Largest diurnal courses were found in UKF, NA, and NLTD. Methods with no diurnal course are the light intercept based methods, NLID, NLIL, and NLI. This was also evident in the synthetic flux analyses (see below). In contrast, methods were very coherent

Table 6 – Ranking of site GPP by each method and ANOVA statistics showing significant differences across sites as classified by all methods

Method/Site	fi1_2001	fi1_2002	it3_2002	Fr4_2002	be1_2000	de3_2001	be1_2001	de3_2000	fr1_2001	fr1_2002
a									■	■
b							■	■		
c						■	■			
d				■	■					
e			■							
f		■								
g	■									
B365	1	2	6	4	3	7	5	8	9	10
NA	1	2	3	5	4	6	7	8	9	10
NE	1	2	3	5	4	7	6	8	9	10
NFA	1	2	3	4	5	6	7	8	9	10
NFO	1	2	4	3	5	6	7	8	9	10
NLID	1	2	4	3	5	7	6	8	10	9
NLIL	1	2	4	3	5	7	8	6	9	10
NLI	1	2	4	3	7	5	8	6	9	10
NLTD	1	2	3	5	6	4	8	7	9	10
NLTL	1	2	3	4	6	5	8	7	9	10
NLT	1	2	3	4	6	5	8	7	9	10
NLSD	1	2	3	6	4	5	7	8	9	10
NLSL	1	2	3	5	4	6	8	7	9	10
NLS	1	2	3	5	4	6	8	7	9	10
SPM	1	2	3	4	5	6	7	8	9	10
UKF	1	2	6	7	3	5	4	8	9	10
NC1	1	2	3	5	4	6	7	8	9	10
NC2	1	2	3	4	5	7	6	8	9	10
MR1	1	2	3	4	5	6	7	8	9	10
MR1R	1	2	3	4	5	6	7	8	9	10
MR2	1	2	3	4	5	6	7	8	9	10
MR2R	1	2	3	4	5	6	7	8	9	10
ANN	1	2	3	4	5	7	6	8	9	10
Max	1	2	6	7	7	7	8	8	10	10
Min	1	2	3	3	3	4	4	6	9	9

Sites that do not share a black box had significantly different GPP from other sites according to the partitioning methods. This analysis indicates that robust comparison across sites is possible given the strong correspondence in site rankings.

with minimal variability on the diurnal pattern of GPP (Fig. 8), which could be expected given the strong direct correlation of photosynthetic active radiation to GPP. Methods were consistent in showing afternoon GPP dip in IT3_2002 and an asymmetric GPP pattern in FR4_2002.

4. Discussion

4.1. *Biases and correlations in model–model comparison*

Retrieval of model generated GPP and RE from noisy modeled NEE data was shown to be feasible for all methods at least on greater than daily timescales. For this analysis, the BETHY model is assumed to be true and thus our results do not necessarily show which methods are more reliable than others, only which methods are better able to decompose a given NEE signal into its components for a given functional form. This is why the B365 method had the highest correlation to the synthetic GPP and RE, due to the similarity in model equations. However, B365 also exhibited a large bias in its retrieval (which can happen because it is a blind parameter retrieval against the noisy model data), showing the need for careful consideration of how using Gaussian cost functions for parameter retrieval may perform poorly in face of non-Gaussian noise.

Most methods were low biased against the synthetic GPP and RE, including the original model itself, on both seasonal and annual scales. The partitioning methods were not biased when comparing method NEE to BETHY NEE, however. Some of this may have been due to the non-Gaussian noise found in eddy covariance flux data and added to the synthetic NEE (Hollinger and Richardson, 2005; Richardson et al., 2006b). The low bias even persisted in statistically sophisticated methods, such as ANN and ANN_S. Alternatively, the BETHY model functions may have forms that do not easily collapse to simple empirical functions used by most methods. Even though the B365 method is based on the BETHY model itself, the MCMC inversions find other parameters with a higher correlation to noisy data at the half-hourly timescale. These parameters, however, lead to the wrong annual GPP and RE. Trudinger et al. (2007) have demonstrated that this failure to retrieve the

Please cite this article in press as: Desai, A.R. et al., Cross-site evaluation of eddy covariance GPP and RE decomposition techniques, Agric. Forest Meteorol. (2007), doi:10.1016/j.agrformet.2007.11.012

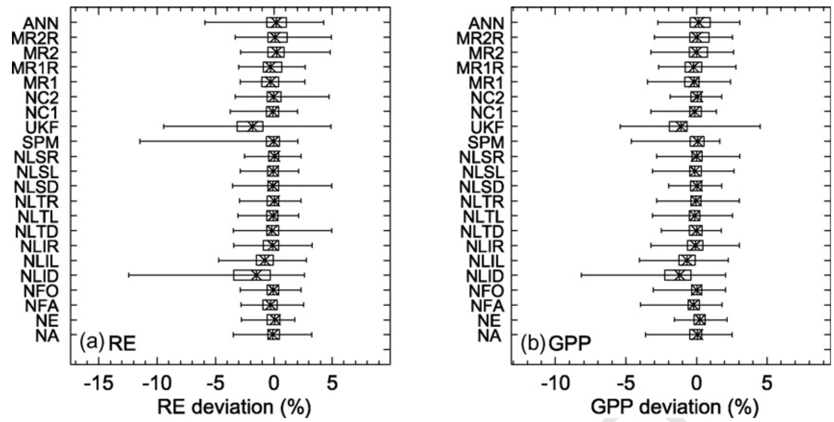

Fig. 4 – Median (star), interquartile range (box) and total range (line) relative sensitivity of methods to 10 mixed gap scenarios averaged across all 10 site years for annual (a) RE and (b) GPP. Most methods did not incur a bias due to gaps, but 10% additional artificial data gaps added on average an increased uncertainty of 8% for RE and 6% for GPP. B365 was excluded since it did not run gap scenarios.

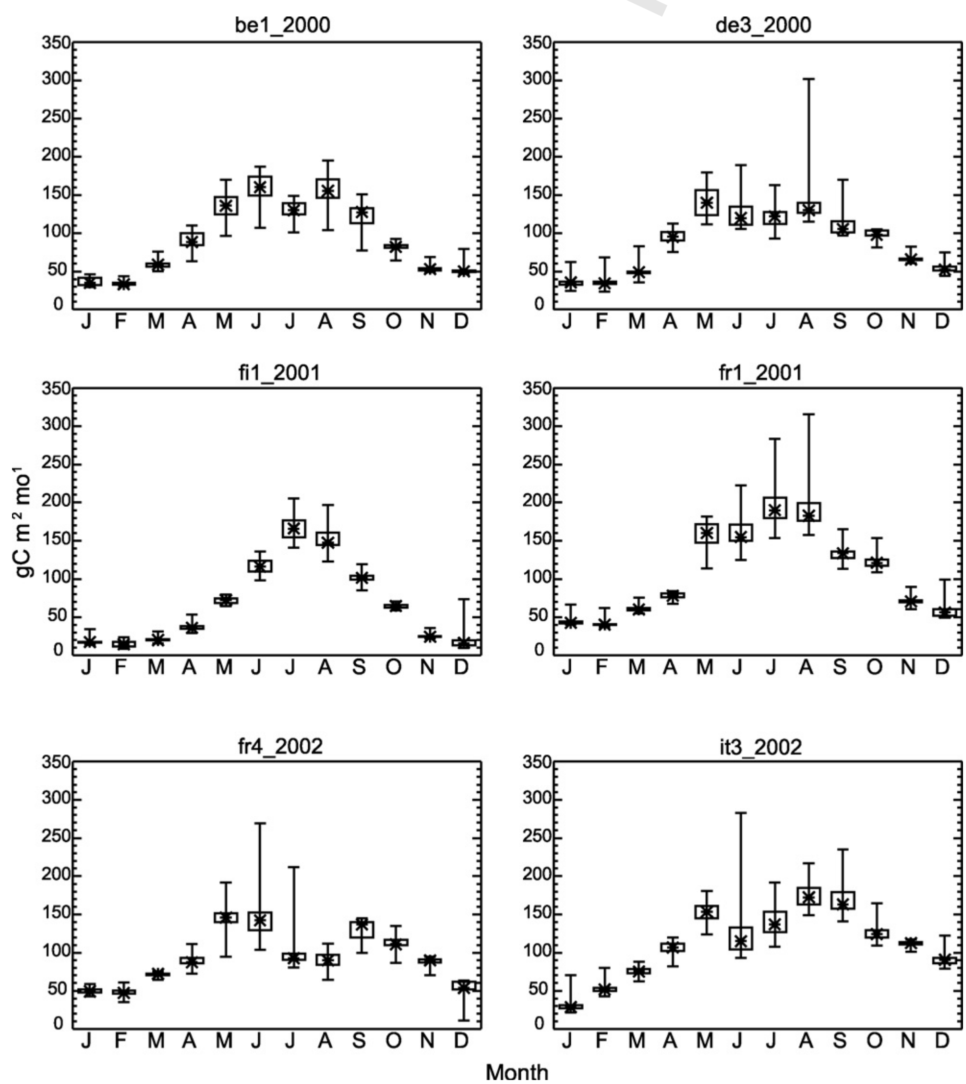

Fig. 5 – Monthly median (star), interquartile range (box) and total range (line) RE as a function of method for a sample year from each of the six unique sites in this study. Generally good agreement was found across most methods on seasonal patterns, though large outliers existed, especially for the Mediterranean sites (lower row).

Please cite this article in press as: Desai, A.R. et al., Cross-site evaluation of eddy covariance GPP and RE decomposition techniques, Agric. Forest Meteorol. (2007), doi:10.1016/j.agrformet.2007.11.012

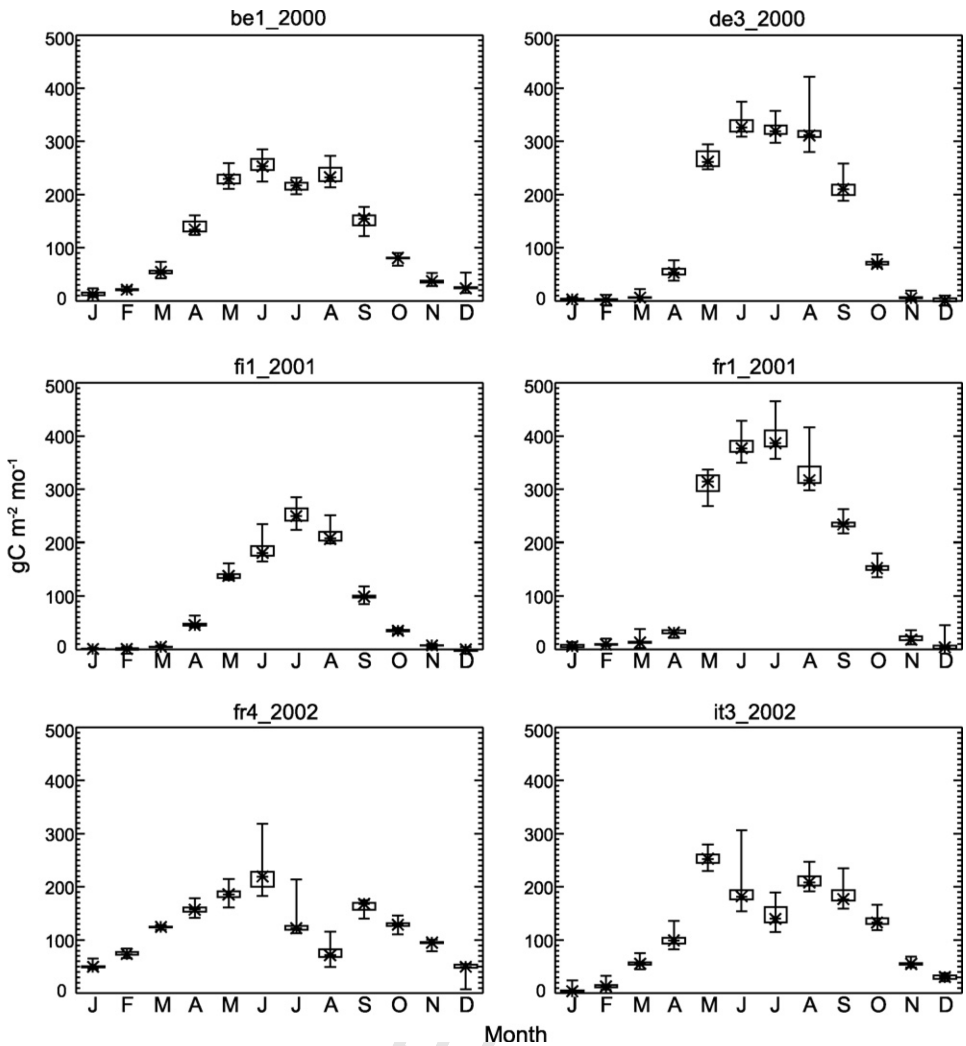

Fig. 6 – Monthly median (star), interquartile range (box) and total range (line) GPP as a function of method for a sample year from each of the six unique sites in this study. The agreement among methods for monthly GPP was stronger than for RE and outliers were smaller.

original BETHY fluxes may well be caused by inconsistencies between the added errors and the cost function used within the optimization. While non-Gaussian errors were added to produce noisy NEE data, B365 applied a Gaussian error model for parameter estimation. The mismatch is in the same range as the overall error of other methods. This highlights the importance of an adequate cost function within the inversion against eddy covariance data. Further research in this direction is needed.

The synthetic analysis did reveal that many methods had low correlation to synthetic RE at hourly timescales. This effect is likely similar to the large variability seen in the diurnal RE in the site diurnal trend analysis. However, the synthetic analysis cannot say that those methods with low correlation are poor at reproducing diurnal RE (though some produce no diurnal signal at all), only poor at recovering the modeled diurnal RE. A significant factor in patterns of diurnal RE is how the partitioning methods incorporate information about air and soil temperature, the latter typically having a damped, lagged signal of air temperature that varies with depth. Given the strong correlations with both temperature variables to RE, diurnal RE patterns from the partitioning methods will generally mimic patterns found in these temperature variables or some combination thereof. All this synthetic analysis can say is whether a method has a diurnal RE pattern similar to BETHY. At daily scales, the previously low correlated methods had large improvement in performance. Multi timescale correlation analysis reveals that most methods except for NLS reach >0.6 correlation to synthetic RE at 8 h averaging time (Fig. 9). The NLS method does not reach that status point until the weekly timescale.

For GPP, high correlation was found at the hourly scale, which increased with averaging time. A small dip was found for all methods except B365 at 12 h. This dip is not easily explained, but should be noted that it is very small and possibly an artifact of BETHY itself, given the model–model nature of the comparison. In both cases, clusters of methods with similar performance metrics do appear, primarily as a function of how closely the methods' functional forms approximate BETHY model's functional forms. Interestingly,

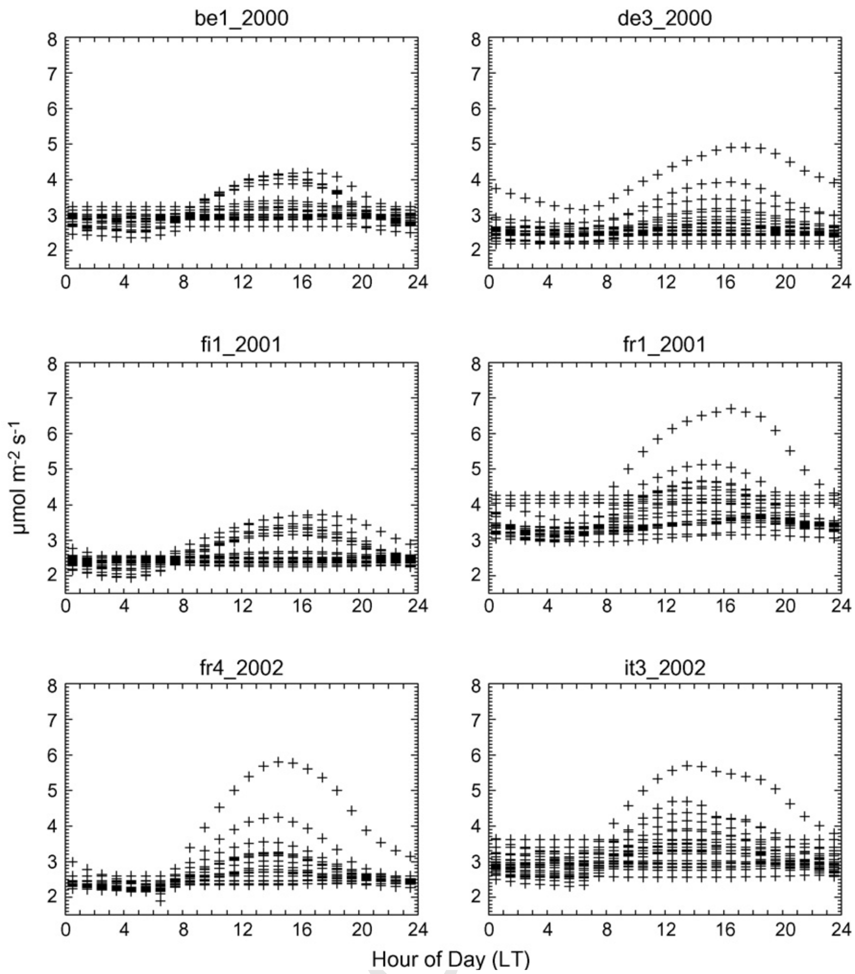

Fig. 7 – Summer (day of year 152–243) ensemble hourly RE for all methods at the six unique sites. Large variation in diurnal course was found across methods partly as a function of relying mainly of air temperature or soil temperature for controlling decomposition.

methods that make few assumptions on seasonal and diurnal patterns, such as ANN, ANN_S and SPM were not leaders in either short timescale correlation or annual bias; rather, many of the non-linear regression methods outperformed them in both metrics. This result suggests the need for more investigation of the newer partitioning methods. Additionally, it should be noted that BETHY is one of many ecosystem models and thus the analysis here should not be construed as a ranking of partitioning methods. BETHY is not necessarily the most complex and complete ecosystem model, but one that represents a broad swatch of these models. Many assumptions need to be made in ecosystem models on the environmental controls of carbon metabolism that do not have strong empirical grounding. The synthetic analysis was performed primarily as an initial way to test variability in GPP and RE retrieved by the NEE partitioning methods for a given known GPP and RE. A more thorough test would be to use an ensemble of models against an ensemble of synthetic noisy scenarios and is recommended here.

The analysis was unable to assess which method was the best for deriving GPP and RE from NEE. Even the synthetic analysis here with one model did not reveal an obvious candidate with both zero bias and high correlation. The analysis did reveal outliers and those with higher variability or bias in the face of gaps, but otherwise we cannot strictly recommend one method over the other. Stoy et al. (2006) compared four methods at three sites with independent data. Though all models performed poorly at estimating short-term RE, they reasoned that their most complex models (non-rectangular hyperbola) that relied on daytime flux data to estimate RE with short time windows, worked best at capturing long timescale variability. Here, we instead find the nighttime extrapolation using short-term temperature sensitivity seemed have highest coincidence with the synthetic data.

4.2. *Total variability in partitioned observed NEE*

Results of the present study demonstrate that multi-site comparisons of component fluxes of NEE, i.e., partitioned GPP and RE, are not valid unless the method used for the partitioning is taken into account. While some methods led to a more or less similar partitioning of NEE, the range across all methods was relatively large (~100 g C m^{-2} $year^{-1}$ IQR), and

Please cite this article in press as: Desai, A.R. et al., Cross-site evaluation of eddy covariance GPP and RE decomposition techniques, Agric. Forest Meteorol. (2007), doi:10.1016/j.agrformet.2007.11.012

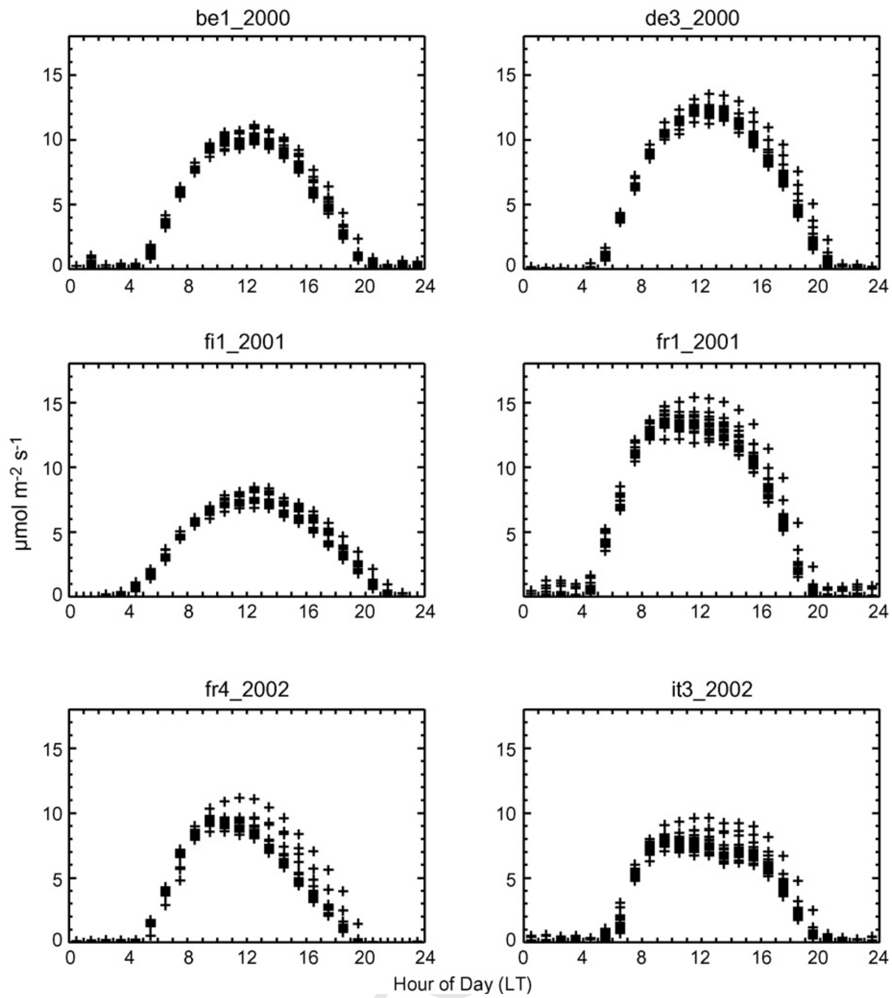

Fig. 8 – Summer (day of year 152–243) ensemble hourly GPP for all methods at the six unique sites. Unlike RE, strong correspondence in diurnal course of GPP was found, due to its strong link to incoming shortwave radiation. Patterns in timing of peak GPP and afternoon GPP decline were evident at some sites, suggesting that studies of environmental controls on photosynthesis are possible with these methods.

this variability may confound true differences among sites. Though this result does not provide an independent confirmation on the fidelity of using eddy flux tower observations for GPP and RE, it does lead to confidence and provide a rough uncertainty bound on previously reported GPP and RE estimates independent of choice of partitioning method. Previous studies focused on a few methods at a many sites (e.g., Falge et al., 2001; Law et al., 2002; Reichstein et al., 2005; Stoy et al., 2006) and so were limited in their ability to draw the conclusions regarding method-related variability in estimated GPP and RE. A few site-specific studies have attempted to use Monte Carlo techniques to evaluate the effects of gaps on both integrated NEE as well as GPP and RE estimates (e.g., Desai et al., 2005; Griffis et al., 2003; Richardson and Hollinger, 2005), but this study is the first to systematically investigate the effects of synthetic gaps on the consistency of the estimated GPP and RE for a range of different partitioning methods.

In this study, across 10 site years of data and 23 methods, 75% of methods fell within 10%, or roughly 100 g C m^{-2} $year^{-1}$, of each other (for a given site year) in terms of annual GPP and RE. Although some outliers were evident at many sites, these were not consistently associated with a particular method, except that for virtually all site years, UKF consistently produced the highest estimates of GPP and RE. The other methods could be separated into groups of models with similar predictions, but no systematic methodological reason can be identified for why some methods fall into one group or the other. Greater variability found for the Mediterranean sites suggests a lack of consensus for partitioning NEE to GPP and RE in seasonally water-limited ecosystems. Hollinger and Richardson (2005) demonstrate that good partitioning methods are approaching the uncertainty limits of the flux data, so larger variability does not necessarily signify poor model selection, but rather that *all* methods are not necessarily suitable for use at *all* kinds of sites depending on core assumptions about seasonal cycles or expected patterns of GPP and RE. In this sense, methods like ANN and SPM, which do not impose *a priori* assumptions about functional relationships between GPP or RE and environmental drivers, may have superior

Please cite this article in press as: Desai, A.R. et al., Cross-site evaluation of eddy covariance GPP and RE decomposition techniques, Agric. Forest Meteorol. (2007), doi:10.1016/j.agrformet.2007.11.012

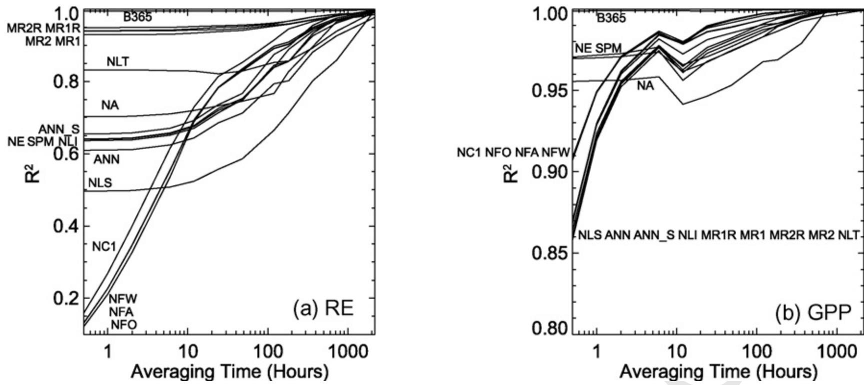

Fig. 9 – Expansion of correlation analysis in Fig. 2 showing correlation as a function of average time for each method compared to the synthetic BETHY model (a) RE and (b) GPP. For RE, all methods except NLS reached R^2 > 0.6 by 12 h despite starting for a wide range of correlation at the half-hourly scale. For GPP, a small dip in correlation was found for all methods except B365 at 12 h, an effect which has not been explained.

performance across a wider range of ecosystem types than empirical regression based routines. Ultimately, though, all partitioning methods will be driven primarily by the variability seen in the driver data provided and if the driver data does not reflect the cause of variation in GPP and RE (e.g., invasive pest outbreak, disturbance, nutrient limitation), then the no method will capture the variation in GPP and RE.

An encouraging aspect of the NEE partitioning methods was their general robustness against artificial data gaps in NEE. Data gaps in flux tower time series are common for a number of reasons and filtering of improper observation conditions will always lead to gaps with eddy covariance. For most methods and sites, 10% additional gaps increased variability of GPP and RE at 75% of sites by 1–2%, but across all sites and methods, variability averaged 6–7%. While these numbers were smaller than the variability caused by choice of partitioning method, it is not an insignificant source of uncertainty. Also, timing and length of gaps matter (e.g., missing a strong respiration peak in early spring), which deserves closer examination (Richardson and Hollinger, ~~in press~~).

Additional variability GPP and RE estimated from flux tower measurements of NEE arrives from systematic corrections to the NEE data such as the u* correction and data filtering, that were not considered in this article (all datasets were already screened for ''bad'' data). Papale et al. (2006) estimate these corrections have an uncertainty less than 100 gC m^{-2} $year^{-1}$ to NEE, leading to potential for ~10% additional uncertainty on GPP and RE estimates. Hagen et al. (2006) used a bootstrapping approach at a single site to estimate uncertainty in GPP due to random errors in eddy covariance data, gaps and GPP model choice. This error turned out to be large at hourly timescales but approached 10% at annual timescales, the largest effect being choice of partitioning method. If all sources of GPP and RE uncertainty assessed here (data filtering (10%), partitioning method choice (10%) and gaps (5%)) were independent and uncorrelated, total uncertainty would on average reach ~25%, limiting the usefulness of comparing GPP and RE, unless they are computed using the same method.

4.3. *Confidence in seasonal and cross-site patterns*

Partitioning methods generally agreed on the cross-site rankings of GPP and RE. These differences were significant according to ANOVA. Intersite ranking of GPP and RE was insensitive to choice of method as long as the same method (or one that is statistically similar) was used for all sites, or the effect of method was considered (e.g., biases are taken into account). The upshot of these results is increased confidence in previously reported comparisons of flux tower derived GPP and RE across sites (e.g., Law et al., 2002; Reichstein et al., 2005), as they should not be strongly affected by choice of method in decomposing the GPP and RE, at least according to this analysis. However, given the variability and biases discussed, comparisons of GPP and RE across sites using different methods are unlikely to have the same coherence, which calls for standardized processing.

Methods were also generally coherent on seasonal trends in GPP and RE at most sites. The ensemble of methods showed close agreement on periods of high and low GPP or RE. However, outliers at a few sites at some months were evident and larger spread was found in the Mediterranean sites, since these sites have seasonal patterns that may not be represented by all methods. Moreover, in the case of IT3_2002, large (multiple week) gaps due to instrumentation issues led to higher uncertainty. Many outliers in other sites were due to one method, typically UKF. Overall, given strong coherence across methods lends support to prior results on studies of GPP and RE seasonality (Falge et al., 2002) and environmental controls on GPP and RE (Law et al., 2002).

Diurnal trends in GPP were coherent across all methods for all sites, which could be expected given the strong and direct correlation between incoming solar radiation and GPP. However, trends for diurnal RE were highly variable across sites, partly driven by method choice of soil temperature or air temperature as the primary control on respiration. Also, mechanisms of diurnal variation for RE are less well known and the timescales on which temperature exhibits control on RE are not well constrained. The filtering of large amounts of nighttime NEE data, existence of inherent noise in flux tower

Please cite this article in press as: Desai, A.R. et al., Cross-site evaluation of eddy covariance GPP and RE decomposition techniques, Agric. Forest Meteorol. (2007), doi:10.1016/j.agrformet.2007.11.012

UNCORRECTED PROOF

time series, and a lack of strong diurnal temperature trend at night places additional limits on the ability of methods to extrapolate diurnal RE from flux tower NEE. In a study using 19 respiration models and data from three flux tower sites, Richardson et al. (2006a) found that neural networks, with their ability to integrate information from multiple forcing and covariance among forcing, performed better than simple parameterized regression models (e.g., Q10, Lloyd-Taylor). However, the focus of that comparison was mainly on annual sums, not diurnal trends. Emerging datasets from automated soil chambers should help quantify actual diurnal trends in soil respiration, which accounts for 40–60% of RE in forested ecosystems (Davidson et al., 2006). It should also be noted that most partitioning methods were designed to characterize the mean but not the variance (or higher order moments) at short timescales, with the intention of producing credible annual means and seasonal cycles rather than preserving all statistics of the time series. Therefore, method performance at the annual timescale should not be taken as a sufficient proxy for performance at short timescales (e.g., diurnal to synpotic) (e.g., Figs. 2 and 9).

5. Summary and conclusions

GPP and RE values estimated by 23 gap-filling methods from 10 site years of NEE flux tower data showed good agreement among methods at the annual and seasonal scales, with variability among methods ~10% of the annual component flux, roughly comparable to typical interannual variability. Artificial gap scenarios (10% data removal) resulted in an additional 6–7% variability for individual methods, but did not tend to bias the method GPP and RE. Most methods were coherent in their ranking of sites from smallest to largest GPP or RE, leading to greater confidence in the ability of these methods to identify cross-site differences and spatial patterns of GPP and RE, *as long as the same method is used to partition NEE across all sites*. In an analysis of synthetic data, we found daily and annual GPP and RE estimates extracted from NEE produced by the BETHY model were generally well correlated with the original synthetic fluxes.

However, there were some notable discrepancies among the partitioning methods. Large outliers existed for some sites and uncertainty was larger for Mediterranean sites. Several of the methods were shown to be systematically biased against the ensemble mean GPP and RE. At the diurnal scale, methods were in close agreement for growing season diurnal GPP course, but varied widely for RE due to choice of functional forms and difficulties in extrapolating high gap frequency nighttime NEE to half-hourly RE. However, no particular class of methods could be identified for having consistent biases. ANOVA analysis did show several individual methods that tended to be biased against the ensemble mean.

As previously stated, this analysis does not identify which methods are more correct in their interpretation of hourly, seasonal or annual GPP and RE. Rather, the results showed the robustness of most methods against the consensus GPP and RE for particular sites, gaps in the NEE data, and coherence of cross-site comparisons. Additionally, the synthetic NEE tests revealed the fidelity of method GPP and RE retrieval, at least for correlation of synthetic to partitioned flux and similarity of the method empirical functions to a complex, well-tested ecosystem model.

Given the relatively fast run times for most methods, the concept of an ensemble modeling system for GPP and RE encompassing different types of methods (data vs. process based; day vs. nighttime based), that were known not to be systematically biased or have large uncertainty/biases with gaps should be explored. Future intercomparison work should focus on comparing methods to independent GPP and RE estimates for the sites, especially long-term automated continuous respiration measurements, which will help with at least the soil respiration component, the source of most ecosystem respiration in many ecosystems. This study showed that additional investigation of the differences of partitioning method results in seasonally water-limited ecosystems, such as the Mediterranean sites, may be needed to better capture GPP and RE. This study only focused on annual data and did not delve specifically into interannual variability. Additional analysis with sets of sites with multiple years of data is warranted, especially in light of the need to move from diagnosis to prediction, which is only possible if we understand the environmental controls on GPP and RE at interannual and longer timescales. Finally, continued development of tests of method fidelity against eddy covariance noise, data filtering, gaps and systematic biases will help further constrain the total expected uncertainty for GPP and RE estimates.

Uncited references

Q2

Aubinet et al. (2000) and Braswell et al. (2005).

Acknowledgements

We wish to acknowledge the site PIs Marc Aubinet (Vielsalm), Werner Kutsch (Hainich), André Granier (Hesse), Serge Rambal (Puechabon), Riccardo Valentini (Roccarespampani) and Timo Vesala (Hyytiala) for making their data available. Data have been collected in the context of Carboeuroflux and CarboeuropeIP research projects funded by the European Commission and part of the sites are also co-funded by local agencies. David Y. Hollinger and Andrew D. Richardson gratefully acknowledge support from the Office of Science (BER), U.S. Department of Energy, Interagency Agreement No. DE-AI02-00ER63028. Ankur R. Desai acknowledges support from the National Science Foundation (NSF), National Center for Atmospheric Research (NCAR) Advanced Study Program (ASP) Fellowship.

REFERENCES

~~Aubinet, M., Grelle, A., Ibrom, A., Rannik, Ü., Moncrieff, J., Foken, T., Kowalski, A., Martin, P.H., Berbigier, P., Bernhofer, C., Clement, R., Elbers, J.A., Granier, A., Grunwald, T.,~~

Please cite this article in press as: Desai, A.R. et al., Cross-site evaluation of eddy covariance GPP and RE decomposition techniques, Agric. Forest Meteorol. (2007), doi:10.1016/j.agrformet.2007.11.012

~~Morgenstern, K., Pilegaard, K., Rebmann, C., Snijders, W., Valentini, R., Vesala, T., 2000. Estimates of the annual net carbon and water exchange of forest: the Euroflux methodology. Adv. Ecol. Res. 30, 112–175.~~

Aubinet, M., Chermanne, B., Vandenhaute, M., Longdoz, B., Yernaux, M., Laitat, E., 2001. Long term carbon dioxide exchange above a mixed forest in the Belgian Ardennes. Agric. Forest Meteorol. 108, 293–315.

Baldocchi, D.D., 2003. Assessing ecosystem carbon balance: problems and prospects of the eddy covariance technique. Global Change Biol. 9, 479–492.

Barr, A.G., Black, T.A., Hogg, E.H., Kljun, N., Morgenstern, K., Nesic, Z., 2004. Inter-annual variability in the leaf area index of a boreal aspen-hazelnut forest in relation to net ecosystem production. Agric. Forest Meteorol. 126, 237–255.

~~Braswell, B.H., Sacks, W.J., Linder, E., Schimel, D.S., 2005. Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations. Global Change Biol. 11, 335–355.~~

Davidson, E.A., Richardson, A.D., Savage, K.E., Hollinger, D.Y., 2006. A distinct seasonal pattern of the ratio of soil respiration to total ecosystem respiration in a spruce-dominated forest. Global Change Biol. 12, 230–239.

Desai, A.R., Bolstad, P., Cook, B.D., Davis, K.J., Carey, E.V., 2005. Comparing net ecosystem exchange of carbon dioxide between an old-growth and mature forest in the upper Midwest, USA. Agric. Forest Meteorol. 128, 33–55.

Falge, E., Baldocchi, D., Olson, R.J., Anthoni, P., Aubinet, M., Bernhofer, C., Burba, G., Ceulemans, R., Clement, R., Dolman, H., Granier, A., Gross, P., Grünwald, T., Hollinger, D., Jensen, N.-O., Katul, G., Keronen, P., Kowalski, A., Lai, C., Law, B.E., Meyers, T., Moncrieff, J., Moors, E., Munger, J.W., Pilegaard, K., Rannik, Ü., Rebmann, C., Suyker, A., Tenhunen, J., Tu, K., Verma, S., Vesala, T., Wilson, K., Wofsy, S., 2001. Gap filling strategies for defensible annual sums of net ecosystem exchange. Agric. Forest Meteorol. 107, 43–69.

Falge, E., Baldocchi, D., Tenhunen, J., Aubinet, M., Bakwin, P., Berbigier, P., Bernhofer, Ch., Burba, G., Clement, R., Davis, K.J., Elbers, J.A., Goldstein, A.H., Grelle, A., Granier, A., GuÃ°mundsson, J., Hollinger, D., Kowalski, A., Katul, G., Law, B., Malhi, Y., Meyers, T., Monson, R., Munger, J.W., Oechel, W., Paw, U.K.T., Pilegaard, K., Rannik, U., Rebmann, C., Suyker, A., Valentini, R., Wilson, K., Wofsy, S., 2002. Seasonality of ecosystem respiration and gross primary production as derived from FLUXNET measurements. Agric. Forest Meteorol. 113, 53–74.

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

Granier, A., Ceschia, E., Damesin, C., Dufrêne, E., Epron, D., Gross, P., Lebaube, S., Le Dantec, V., Le Goff, N., Lemoine, D., Lucot, E., Ottorini, J.M., Pontailler, J.Y., Saugier, B., 2000. The carbon balance of a young Beech forest. Funct. Ecol. 14, 312–325.

Griffis, T.J., Black, T.A., Morgenstern, K., Barr, A.G., Nesic, Z., Drewitt, G.B., Gaumont-Guay, D., McCaughey, J.H., 2003. Ecophysiological controls on the carbon balances of three southern boreal forests. Agric. Forest Meteorol. 117, 53–71.

Hagen, S.C., Braswell, B.H., Linder, E., Frolking, S., Richardson, A.D., Hollinger, D.Y., 2006. Statistical uncertainty of eddy flux-based estimates of gross ecosystem carbon exchange at Howland Forest, Maine. J. Geophys. Res. 111 (D08S03), doi:10.1029/2005JD006154.

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

Knohl, A., Schulze, E.-D., Kolle, O., Buchmann, N., 2003. Large carbon uptake by an unmanaged 250-year-old deciduous forest in Central Germany. Agric. Forest Meteorol. 118, 151–167.

Knorr, W., Kattge, J., 2005. Inversion of terrestrial ecosystem model parameter values against eddy covariance measurements by Monte Carlo sampling. Global Change Biol. 11, 1333–1351.

Law, B.E., Falge, E., Gu, L., Baldocchi, D., Bakwin, P., Berbigier, P., Davis, K.J., Dolman, H., Falk, M., Fuentes, J., Goldstein, A.H., Granier, A., Grelle, A., Hollinger, D., Janssens, I., Jarvis, P., Jensen, N.O., Katul, G., Malhi, Y., Matteucci, G., Monson, R., Munger, J.W., Oechel, W., Olson, R., Pilegaard, K., Paw, U.K.T., Thorgeirsson, H., Valentini, R., Verma, S., Vesala, T., Wilson, K., Wofsy, S., 2002. Carbon dioxide and water vapor exchange of terrestrial vegetation in response to environment. Agric. Forest Meteorol. 113, 97–120.

~~Moffat, A.M., Papale, D., Reichstein, M., Barr, A.G., Braswell, B., Churkina, G., Desai, A.R., Falge, E., Gove, J.H., Heimann, M., Hollinger, D.Y., Hui, D., Jarvis, A.J., Kattge, J., Noormets, A., Richardson, A.D., Stauch, V.J., in review. Comprehensive comparison of gap filling techniques for net carbon fluxes. Agric. Forest Meteorol., in press, doi:10.1016/j.agrformet.2007.08.011.~~ Q3

Noormets, A., Chen, J., Crow, T.R. Age-dependent changes in ecosystem carbon fluxes in managed forests in northern Wisconsin, USA. Ecosystems, ~~in press~~. Q4

Papale, D., Valentini, R., 2003. A new assessment of European forests carbon exchanges by eddy fluxes and artificial neural network spatialization. Global Change Biol. 9, 525–535.

Papale, D., Reichstein, M., Aubinet, M., Canfora, E., Bernhofer, C., Longdoz, B., Kutsch, W., Rambal, S., Valentini, R., Vesala, T., Yakir, D., 2006. Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: algorithms and uncertainty estimation. Biogeosciences 3, 571–583.

Peylin, P., Bousquet, P., Le Quéré, C., Sitch, S., Friedlingstein, P., McKinley, G., Gruber, N., Rayner, P., Ciais, P., 2005. Multiple constraints on regional CO_2 flux variations over land and oceans. Global Biogeochem. Cycles 19 (GB1011), doi:10.1029/2003GB002214.

Rambal, S., Joffre, R., Ourcival, J.M., Cavender-Bares, J., Rocheteau, A., 2004. The growth respiration component in eddy CO_2 flux from a Quercus ilex mediterranean forest. Global Change Biol. 10, 1460–1469.

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grunwald, T., Havrankova, K., Ilvesniemi, H., Janous, D., Knohl, A., Laurila, T., Lohila, A., Loustau, D., Matteucci, G., Meyers, T., Miglietta, F., Ourcival, J.M., Pumpanen, J., Rambal, S., Rotenberg, E., Sanz, M., Tenhunen, J., Seufert, G., Vaccari, F., Vesala, T., Yakir, D., Valentini, R., 2005. On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. Global Change Biol. 11, 1424–1439.

Richardson, A.D., Hollinger, D.Y., 2005. 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. Forest Meteorol. 131, 191–208.

Richardson, A.D., Braswell, B.H., Hollinger, D.Y., Burman, P., Davidson, E.A., Evans, R.S., Flanagan, L.B., Munger, J.W.,

UNCORRECTED PROOF

Savage, K., Urbanski, S.P., Wofsy, S.C., 2006a. Comparing simple respiration models for eddy flux and dynamic chamber data. Agric. Forest Meteorol. 141, 219–234.

Richardson, A.D., Hollinger, D.Y., Davis, K.J., Flanagan, L.B., Katul, G.G., Stoy, P.C., Verma, S.B., Wofsy, S.C., 2006b. A multi-site analysis of uncertainty in tower-based measurements of carbon and energy fluxes. Agric. Forest Meteorol. 136, 1–18.

Richardson, A.D., Holinger, D.Y., Aber, J.D., Ollinger, S.V., Braswell, B.H., 2007. Environmental variation is directly responsible for short- but not long-term variation in forest-atmosphere carbon exchange. Global Change Biol. 13, 788–803, doi:10.1111/j.1365-2486.2007.01330.x.

Richardson, A.D., Hollinger, D.Y. The addition uncertainty in gap filled NEE results from long gaps in the CO_2 flux record. Q5 Agric. Forest Meteorol., in press.

Stauch, V.J., Jarvis, A.J., 2006. A semi-parametric gap-filling model for eddy covariance CO_2 flux time series data. Global Change Biol. 12, 1707–1716.

Stauch, V.J., 2007. Data-led methods for the analysis and interpretation of eddy covariance CO_2 flux observations. Ph.D. Thesis. University of Potsdam. http://opus.kobv.de/ubp/volltexte/2007/1238/pdf/stauch_diss.pdf.

Stoy, P.C., Katul, G.G., Siqueira, M.B.S., Juang, J.-Y., Novick, K.A., Oren, R., 2006. An evaluation of methods for partitioning eddy covariance-measured net ecosystem exchange into photosynthesis and respiration. Agric. Forest Meteorol. 141, 2–18.

Suni, T., Rinne, J., Reissell, A., Altimir, N., Keronen, P., Rannik, Ü., Dal Maso, M., Kulmala, M., Vesala, T., 2003. Long-term measurements of surface fluxes above a Scots pine forest in Hyytiala, southern Finland, 1996-2001. Boreal Environ. Res. 8, 287–301.

Tedeschi, V., Rey, A.N.A., Manca, G., Valentini, R., Jarvis, P.G., Borghetti, M., 2006. Soil respiration in a Mediterranean oak forest at different developmental stages after coppicing. Global Change Biol. 12, 110–121.

Trudinger, C.M., Raupach, M.R., Rayner, P.J., Kattge, J., Liu, Q., Pak, B., Reichstein, M., Renzullo, L., Richardson, A.D., Roxburgh, S.H., Styles, J., Wang, Y.-P., Briggs, P., Barrett, D., Nikolova, S., 2007. OptIC project: an intercomparison of optimization techniques f or parameter estimation in terrestrial biogeochemical models. J. Geophys. Res. 112 (G02027), doi:10.1029/2006JG000367.

UNCORRECTED PROOF

Please cite this article in press as: Desai, A.R. et al., Cross-site evaluation of eddy covariance GPP and RE decomposition techniques, Agric. Forest Meteorol. (2007), doi:10.1016/j.agrformet.2007.11.012