Biogeosciences

Open Access

Evaluating the agreement between measurements and models of net ecosystem exchange at different times and timescales using wavelet coherence: an example using data from the North American Carbon Program Site-Level Interim Synthesis

P. C. Stoy[1], M. C. Dietze[2], A. D. Richardson[3], R. Vargas[4], A. G. Barr[5], R. S. Anderson[6], M. A. Arain[7], I. T. Baker[8],
T. A. Black[9], J. M. Chen[10], R. B. Cook[11], C. M. Gough[12], R. F. Grant[13], D. Y. Hollinger[14], R. C. Izaurralde[15],
C. J. Kucharik[16], P. Lafleur[17], B. E. Law[18], S. Liu[19], E. Lokupitiya[20], Y. Luo[21], J. W. Munger[22], C. Peng[23],
B. Poulter[24], D. T. Price[25], D. M. Ricciuto[11], W. J. Riley[26], A. K. Sahoo[27], K. Schaefer[28], C. R. Schwalm[29], H. Tian[30],
H. Verbeeck[31], and E. Weng[32]

[1]Department of Land Resources and Environmental Sciences, Bozeman, MT 59717, USA
[2]Department of Earth and Environment, Boston University, Boston, MA 02215, USA
[3]Department of Organismic & Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA
[4]Department of Plant and Soil Sciences, Delaware Environmental Institute, University of Delaware, Newark, DE 19717, USA
[5]Climate Research Division, Atmospheric Science and Technology Directorate, Saskatoon, SK S7N 3H5, Canada
[6]Numerical Terradynamic Simulation Group, University of Montana, Missoula, MT 59812, USA
[7]School of Geography and Earth Sciences and McMaster Centre for Climate Change, McMaster University, Hamilton, ON
L8S 4K1, Canada
[8]Department of Atmospheric Science, Colorado State University, Fort Collins, CO 80523, USA
[9]Faculty of Land and Food Systems, University of British Columbia, Vancouver, BC V6T 1Z4, Canada
[10]Department of Geography and Program in Planning, University of Toronto, Toronto, ON M5S 3G3, Canada
[11]Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA
[12]Department of Biology, Virginia Commonwealth University, Richmond, VA 23284, USA
[13]Department of Renewable Resources, University of Alberta, Edmonton, AB T6G 2E3, Canada
[14]Northern Research Station, USDA Forest Service, Durham, NH 03824, USA
[15]Pacific Northwest National Laboratory and University of Maryland, College Park, MD 20740, USA
[16]Department of Agronomy & Nelson Institute Center for Sustainability and the Global Environment, University of
Wisconsin – Madison, Madison, WI 53706, USA
[17]Department of Geography, Trent University, Peterborough, ON K9J 7B8, Canada
[18]Department of Forest Ecosystems and Society, Oregon State University, Corvallis, OR 97331, USA
[19]US Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center, Sioux Falls, SD 57198, USA
[20]Department of Zoology, POB 1490, University of Colombo, Colombo 03, Sri Lanka
[21]Department of Botany and Microbiology, University of Oklahoma, Norman, OK 73019, USA
[22]School of Engineering and Applied Sciences and Department of Earth and Planetary Sciences, Harvard University,
Cambridge, MA 02138, USA
[23]Department of Biology Sciences, University of Quebec at Montreal, Montreal, QC H3C 3P8, Canada
[24]Laboratoire des Sciences du Climat et de l'Environnement, Gif-sur-Yvette, France
[25]Northern Forestry Centre, Canadian Forest Service, Edmonton, AB T6H 3S5, Canada
[26]Climate and Carbon Sciences, Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[27]Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ 08544, USA
[28]National Snow and Ice Data Center (NSIDC), University of Colorado, Boulder, CO 80309, USA
[29]School of Earth Science and Environmental Sustainability, Northern Arizona University, Flagstaff, AZ 86001, USA
[30]School of Forestry and Wildlife Sciences, Auburn University, Auburn, AL 36849, USA

[31]Laboratory of Plant Ecology, Ghent University, 9000 Ghent, Belgium
[32]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA

Correspondence to: P. C. Stoy (paul.stoy@montana.edu)

Abstract. Earth system processes exhibit complex patterns across time, as do the models that seek to replicate these processes. Model output may or may not be significantly related to observations at different times and on different frequencies. Conventional model diagnostics provide an aggregate view of model–data agreement, but usually do not identify the time and frequency patterns of model–data disagreement, leaving unclear the steps required to improve model response to environmental drivers that vary on characteristic frequencies. Wavelet coherence can quantify the times and timescales at which two time series, for example time series of models and measurements, are significantly different. We applied wavelet coherence to interpret the predictions of 20 ecosystem models from the North American Carbon Program (NACP) Site-Level Interim Synthesis when confronted with eddy-covariance-measured net ecosystem exchange (NEE) from 10 ecosystems with multiple years of available data. Models were grouped into classes with similar approaches for incorporating phenology, the calculation of NEE, the inclusion of foliar nitrogen (N), and the use of model–data fusion. Models with prescribed, rather than prognostic, phenology often fit NEE observations better on annual to interannual timescales in grassland, wetland and agricultural ecosystems. Models that calculated NEE as net primary productivity (NPP) minus heterotrophic respiration (HR) rather than gross ecosystem productivity (GPP) minus ecosystem respiration (ER) fit better on annual timescales in grassland and wetland ecosystems, but models that calculated NEE as GPP minus ER were superior on monthly to seasonal timescales in two coniferous forests. Models that incorporated foliar nitrogen (N) data were successful at capturing NEE variability on interannual (multiple year) timescales at Howland Forest, Maine. The model that employed a model–data fusion approach often, but not always, resulted in improved fit to data, suggesting that improving model parameterization is important but not the only step for improving model performance. Combined with previous findings, our results suggest that the mechanisms driving daily and annual NEE variability tend to be correctly simulated, but the magnitude of these fluxes is often erroneous, suggesting that model parameterization must be improved. Few NACP models correctly predicted fluxes on seasonal and interannual timescales where spectral energy in NEE observations tends to be low, but where phenological events, multi-year oscillations in climatological drivers, and ecosystem succession are known to be important for determining ecosystem function. Mechanistic improvements to models must be made to replicate observed NEE variability on seasonal and interannual timescales.

1 Introduction

Land surface models represent our understanding of how terrestrial ecosystems function in the climate system. It is critical to test, compare, and improve these models as new information and methods become available, especially because numerous recent syntheses have demonstrated a considerable lack of model skill when confronted with observations (Schwalm et al., 2010; Wang et al., 2010; Schaefer et al., 2012). Models are commonly diagnosed using statistical metrics that can be combined for a more complete view of model performance (Taylor, 2001). Such model diagnostics are able to identify whether a different model, different model parameterization, or different subroutine represents an improvement (Akaike, 1974), but are not intended to identify the symptoms of model–data disagreement across time and scales in time in order to identify the conditions that result in discrepancies. Residual analyses and detailed investigations of model performance during different time periods give important insight into the mechanisms underlying model failure, but are rarely interpreted with respect to patterns of model/measurement mismatch in the frequency (or timescale) domain (see however Dietze et al., 2011; Mahecha et al., 2010; Vargas et al., 2010, and Vargas et al., 2013). In this paper, we quantify periods in time and scales in time when ecosystem models are not significantly related to eddy covariance measurements of net ecosystem exchange (NEE) to identify periods in which models can and should be improved (Williams et al., 2009).

Improving individual models is a noteworthy goal, but modern studies often combine multiple observations and multiple model simulations (i.e., multiple databases) to arrive at a synthesis (Friedlingstein et al., 2006; Schwalm et al., 2010). In other words, such studies adopt a data-intensive approach to scientific inference (Gray, 2009), and techniques from nonlinear time-series analysis and knowledge discovery in databases (i.e., "data mining") may provide important

insights into the aggregate or divergent behavior of these model and observational databases. In this study, we quantify significant relationships among 20 ecosystem models and 10 multi-year time series of NEE measurements from the North American Carbon Program (NACP) Site-Level Interim Synthesis (Schwalm et al., 2010) using a technique called wavelet coherence (Grinsted et al., 2004; Torrence and Webster, 1999). Wavelet coherence is conceptually similar to a measure of correlation between data series across time and timescale (related to frequency). Like correlation, significant values of wavelet coherence can be quantified, in this case by comparison against appropriate synthetic null spectra. Unlike simple correlation, statistical significance can be quantified across both time and timescales simultaneously. We use wavelet coherence to determine the times and timescales when NACP models and measurements are significantly related and, more importantly, when they are not. Notably, wavelet coherence can quantify significance in the time and timescale domains even when common power (i.e., shared variability) among time series on these scales is low (Grinsted et al., 2004), and may offer an improvement over residual analyses for this reason. Wavelet coherence has found applications in comparing ecological models and measurements for the goal of model improvement (Williams et al., 2009; Wang et al., 2011), but not across multiple model and observational time series to date.

Previous studies of ecosystem models in the timescale domain have demonstrated that models tend to miss patterns in flux observations on intermediate (i.e., weekly to monthly) and interannual timescales (Siqueira et al., 2006; Stoy et al., 2005). Biological responses to variability in climate often dominates flux variability on these timescales (Richardson et al., 2007), and models tend to replicate this biological functioning poorly. Such responses include weekly to monthly shifts in leaf-out/leaf-drop phenology and the multitude of factors, including lagged responses, known to contribute to interannual carbon flux variability (Richardson et al., 2013). With respect to the NACP, findings to date have identified superior model fit when phenology is prescribed by remote sensing observations as opposed to prognostic via a phenology model, when a sub-daily (i.e., half hourly or hourly) rather than a daily time step is used, and when net ecosystem exchange (NEE) is calculated as the difference between gross primary productivity (GPP) and ecosystem respiration (ER) rather than the difference between net primary productivity (NPP) and heterotrophic respiration (HR) (Schwalm et al., 2010; Richardson et al., 2012). Schwalm et al. (2010) also found that model performance was poorer during spring and fall when phenological events dominate surface flux and during dry periods within the growing season. Less certain is how models match measurements on multiple timescales as they respond to climatic and biological forcings that act on multiple timescales (Dietze et al., 2011). Quantifying such model–measurement relationships contributes to the NACP objective to measure and understand the sources and sinks of CO_2 in North America. Following previous studies, we hypothesize that models will tend to match flux patterns on daily and annual timescales, and we focus our investigation on timescales between weeks and multiple months as well as interannual timescales, where we postulate that models will replicate observations more poorly.

2 Methods

2.1 Eddy covariance data and ecosystem models

Half hourly (or hourly) micrometeorological and eddy covariance measurements were collected by site principal investigators and research teams and were provided to the AmeriFlux and Fluxnet-Canada consortia to create the NACP Site Level Interim Synthesis product (Schwalm et al., 2010). For this analysis we examine 20 ecosystem models against measurements of the net ecosystem exchange of CO_2 (NEE) from the 10 eddy covariance research sites investigated by Dietze et al. (2011) (Table 1). These sites were chosen because the length of the observation period tended to be longer and more continuous, allowing us to investigate interannual (multiple year) variability, and because more models tended to be run for these ecosystems (Schwalm et al., 2010; Schaefer et al., 2012). Missing meteorological data were gap-filled using National Oceanic and Atmospheric Administration (NOAA) meteorological station data and Daymet reanalysis products (Ricciuto et al., 2009). Half-hourly (or hourly) NEE values were filtered to remove periods of insufficient turbulence determined using friction velocity (u^*) thresholds, and despiked to remove outliers (Papale et al., 2006; Reichstein et al., 2005). Missing NEE data were then gap-filled following Barr et al. (2009). We note that gap-filling models tend to match closely the orthonormal wavelet coefficients of NEE measurements at timescales longer than one day (Stoy et al., 2006), and thus gap-filling artifacts should have minimal impact on our results. Model runs at each site followed a prescribed protocol for intercomparison described by Schwalm et al. (2010). Ancillary biological, disturbance, edaphic, and management data used by model runs for each site were given by the AmeriFlux BADM templates (Law et al., 2008). The ecosystem models explored here are listed in Table 2 and described in more detail in Schwalm et al. (2010) and the original publications. Likewise, information regarding the study sites is best found in the original publications (Table 1).

2.2 Wavelet coherence

The times and timescales at which two corresponding data series (here time series) have high common power can be quantified using the wavelet cospectrum. Wavelet coherence uses wavelet spectral and cospectral calculations to quantify correlations in the time and timescale domains (Grinsted et al., 2004; Torrence and Webster, 1999), and we refer

Table 1. Measurement sites of the North American Carbon Program Site-Level Interim Synthesis investigated by Dietze et al. (2011) and explored in the present analysis. CRO: crop; DBF: deciduous broadleaf forest; ENF: evergreen needleleaf forest; GRA: grassland; PFT: plant functional type; WET: wetland.

Site ID	Name	PFT	Years	Figs. 6–9 subplot	References
CA-Ca1	Campbell River: Mature Forest Site	ENF	1998–2006	A	Schwalm et al. (2007)
CA-Let	Lethbridge	GRA	1999–2007	B	Flanagan et al. (2002)
CA-Mer	Eastern Peatland: Mer Bleue	WET	1999–2006	C	Lafleur et al. (2003)
CA-Oas	SSA Old Aspen	DBF	1997–2006	D	Krishnan et al. (2006)
CA-Obs	SSA Old Black Spruce	ENF	2000–2006	E	Griffis et al. (2003)
US-Ha1	Harvard Forest	DBF	1992–2005	See Figs. 1–5	Urbanski et al. (2007)
US-Ho1	Howland Forest	ENF	1996–2004	F	Richardson et al. (2009)
US-Me2	Metolius: Intermediate Aged Ponderosa Pine	ENF	2002–2007	G	Thomas et al. (2009)
US-Ne3	Mead: Rainfed Maize/Soybean Rotation	CRO	2002–2004	H	Verma et al. (2005)
US-UMB	University of Michigan Biological Station	DBF	1999–2006	I	Schmid et al. (2003), Gough et al. (2008), Gough et al. (2009)

Table 2. A list of model attributes per model following Schwalm et al. (2010). Model/attribute combinations with no checked boxes indicate that a different formulation was used. These are not considered here. ER: ecosystem respiration; GPP: gross primary productivity; HR: heterotrophic respiration; NEE: net ecosystem exchange; NPP: net primary productivity.

Model	NEE calc.		Phenology		Foliar N		Reference
	NPP − HR	GPP − ER	Prognostic	Prescribed	Yes	No	
AgroIBIS	X		X		X		Kucharik and Twine (2008)
BEPS	X		X		X		Liu et al. (1999)
Biome-BGC			X		X		Thornton et al. (2005)
Can-IBIS	X		X		X		Williamson et al. (2008)
CN-CLASS		X	X		X		Arain et al. (2008)
DLEM	X		X[+]		X		Tian et al. (2010)
DNDC	X		X		X		Li et al. (2010)
Ecosys		X	X		X		Grant et al. (2005)
ED2	X		X		X		Medvigy et al. (2009)
EPIC	X		X		X		Causarano et al. (2007)
ISOLSM		X		X		X	Riley et al. (2002)
LoTEC	X		X			X	Hanson et al. (2004)
LPJ-wsl	X		X			X	Sitch et al. (2003)
ORCHIDEE	X		X			X	Krinner et al. (2005)
SiB3		X		X	X[*]		Baker et al. (2008)
SiBCASA		X		X		X	Schaefer et al. (2009)
SiBcrop		X	X		X		Lokupitiya et al. (2009)
SSiB2		X		X		X	Zhan et al. (2003)
TECO		X	X			X	Weng and Luo (2008)
Triplex-FLUX		X		X		X	Zhou et al. (2008)

[+] Semi-prognostic phenology.
[*] SIB3 includes N in the assignment of phenology from remotely sensed products, but does not otherwise include it as a prognostic variable.

the reader to the original publications for a detailed explanation of wavelet coherence. Briefly, following Grinsted et al. (2004), wavelet coherence is defined in a similar manner to the coefficient of determination (R^2) using instead wavelet coefficients. Wavelet coherence, rather than the wavelet cross spectrum, is preferred for significance testing as spurious peaks in the wavelet cross spectrum often result if one, but

not necessarily both, of the time series under investigation exhibits high power at a given time and timescale (Maraun and Kurths, 2004).

Grinsted et al. (2004) noted that many geophysical time series are characterized by red (Brownian) noise, which can be modeled as a first-order autoregressive process (AR1). These patterns can be used as a null model by simulating synthetic

data that were simulated with AR1 coefficients to quantify significant wavelet coherence at the 95 % confidence level. Eddy covariance time series approximate pink noise ($1/f$ noise) (Richardson et al., 2008), which is likewise a class of autoregressive noise, and Grinsted et al. (2004) demonstrated that the color of noise has little impact on the determination of the significance level. Wavelet coherence values above 0.7 were found by Grinsted et al. (2004) to be significant against synthetic data sets across a wide range of scales when ten scales per octave (i.e., per a doubling or halving of frequency) were chosen in the scale-wise smoothing, although higher coherence values (ca. 0.8 or higher) should be chosen at very high and low frequencies. We used ten scales per octave and also chose the commonly used 0.7 wavelet coherence threshold for determining significance. We de-emphasize the interpretation of high-frequency coherence (e.g., on hourly and sub-daily timescales) to focus on the longer timescales (i.e., > one day) where models often fail. Wavelet coefficients on very long timescales (years to multiple years) often exceed the so-called cone of influence beyond which the coherence calculation is dominated by edge effects because of incomplete time locality across frequencies (Torrence and Compo, 1998). Wavelet coefficients outside the cone of influence are unreliable and will not be interpreted here. Also for consistency with Grinsted et al. (2004), we chose the Morlet wavelet basis function with a wave number of six. Time series were truncated to powers of two for spectral calculations. Further, it is common to present wavelet coefficients as the absolute value of their real and imaginary components along time and frequency axes: the so-called wavelet half-plane. Here, we present wavelet coherence values in a similar manner along the time and timescale axes.

The dimensions of time and timescale (subsequently called "regions") adjudged to be significant may occasionally be larger than uncertainty bounds on account of autocorrelation in time and timescale (Maraun et al., 2007). As a consequence, we advocate a conservative analysis of the precise dimensions of regions adjudged to have significant coherence, and we do not seek here to interpret the dimensions of all regions here, rather the existence of significant wavelet coherence. We also note that, following the Monte Carlo analysis, significant wavelet coherence may occur by chance. Our analysis focuses on regions where models and measurements do not exhibit significant coherence under the expectation that models should match measurements, and with the understanding that one learns more about a given system when models fail.

Results are presented with two different representations of timescale in mind. For the demonstration of the wavelet coherence technique, we interpret all relevant scales from twice the observation time step (usually 1 to 2 hours) to half the length of the truncated time series. For the comparison of model output against flux observations, we interpret wavelet coherence on timescales longer than one day to enable a comparison among models that operate on daily and sub-daily time steps and to focus our analysis on the longer timescales (e.g., seasonal or interannual timescales) on which models often fail.

2.2.1 Combined wavelet coherence significance analysis

A time–timescale graph of wavelet coherence significance values can be created for each model–measurement combination for each site. As such, significance values from different models run for a single site that lie upon identical axes can be combined for an aggregate view of model performance. The approach that we explore is to sum wavelet coefficients that represent significance values at different times and timescales (i) for models that possess a given attribute A (A_i), divide by the number of models with A (N_A), then subtract the sum of the wavelet coherence significance values for models that possess the opposing model attribute B (B_i) divided by the number of models with B (N_B):

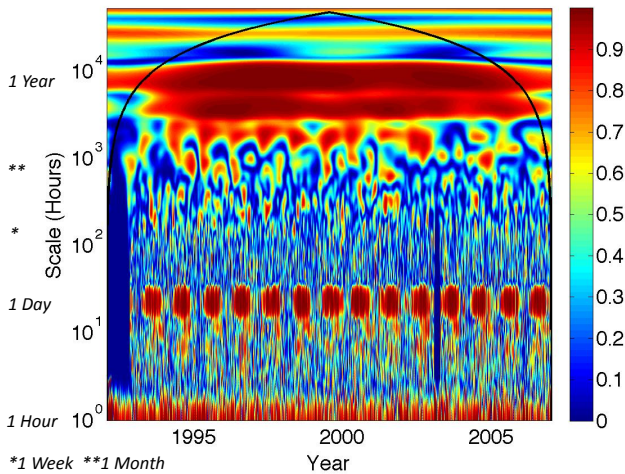$$\frac{1}{N_A} \sum_{i=1}^{N_A} A_i - \frac{1}{N_B} \sum_{i=1}^{N_B} B_i. \tag{1}$$

The purpose of this calculation is to provide a simple metric between -1 and 1 for cases where N_A and N_B may be different but are weighted equally to simplify comparison. The goal is to identify regions in time and timescale at different sites where a certain model attribute outperforms the other (or others) across all models investigated here, with the goal of interpreting the success or failure of different model formulations across time and timescale for different ecosystem types. An infinite number of alternate approaches to compare multiple models exists. To avoid over-interpreting results and to simplify the visual display, we only plot absolute values of Eq. (1) that exceed 0.33 to focus our study on times and timescales where the first and second terms of Eq. (1) differ by at least one-third.

3 Results and discussion

3.1 Wavelet coherence

We begin by demonstrating significance testing using wavelet coherence with a single site/model combination. The time series of NEE from the Harvard Forest (US-Ha1) site encompasses 140 256 potential hourly observations from 1991 until the end of 2006 (Urbanski et al., 2007). We interpret the 2^{17} (=131 072) NEE measurements between 18 January 1992 and the end of the time series and NEE simulations from the Ecosystem Demography model version 2 (ED2, Medvigy et al., 2009). The wavelet coherence between US-Ha1 and ED2 tends to be large (> 0.7) on the daily timescale (24 h, ca. $10^{1.38}$) during growing seasons and on the annual timescale (8760 h, ca. $10^{3.94}$) across the entire measurement period (Fig. 1). Measured NEE from US-Ha1

Fig. 1. Wavelet coherence between Harvard Forest (US-Ha1) net ecosystem exchange of CO_2 (NEE) observations and Ecosystem Demography 2 (ED2) model simulations along the time and scale axes. The model was not run for the first year of measurements for the North American Carbon Program Site-Level Interim Synthesis. The black line is the cone of influence beyond which wavelet coefficients should not be interpreted.
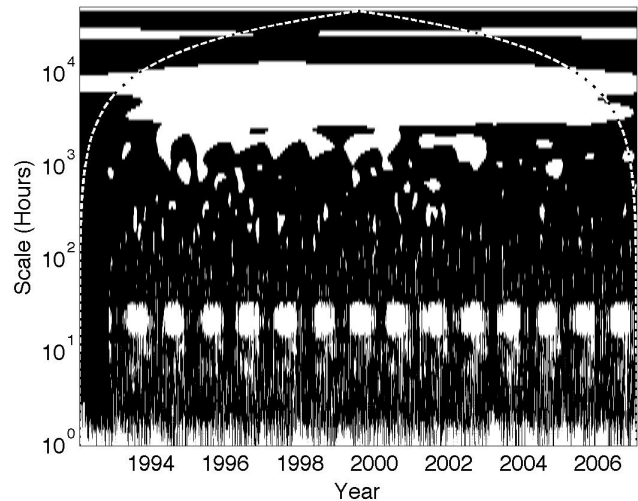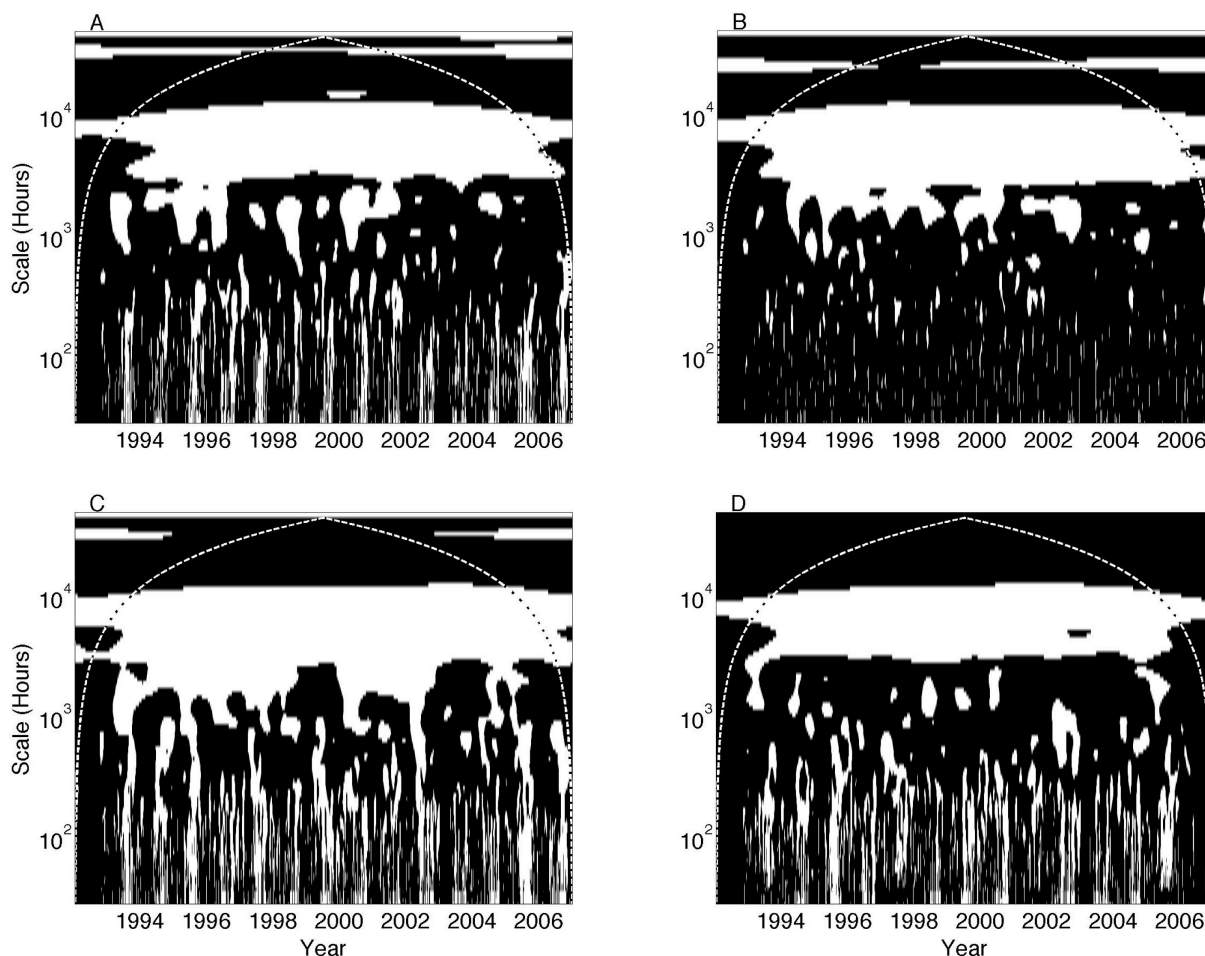

Fig. 2. Same as Fig. 1 but for regions of significant wavelet coherence between net ecosystem exchange (NEE) measurements at Harvard Forest (US-Ha1) and the Ecosystem Demography v2 (ED2) model simulations, calculated following Grinsted et al. (2004). Regions in time and timescale with significant coherence are white and given a value of unity; regions without significant coherence are black and given the null value.

and modeled NEE from ED2 demonstrate common power on these timescales. Some multi-day to multi-month (seasonal) periods likewise have high wavelet coherence, but wavelet coherence is generally low (<0.7) on timescales longer than one year.

3.2 Wavelet coherence significance testing

Wavelet coherence coefficients were converted to binary significance values as demonstrated in Fig. 2. Here, regions in time and timescale that have significant coherence at the 95 % level (i.e., wavelet coherence coefficients > 0.7 following Grinsted et al., 2004) are given the value of one and appear in white in the figure, and non-significant regions are given the value of zero and appear in black in the figure. Figure 2 reveals that ED2-modeled NEE is significantly related to the NEE measurements on daily timescales during the growing season (i.e., the white areas in Fig. 2), on the annual timescale, and on seasonal timescales during the earlier part of the measurement period, but not during most of the remaining times and timescales. Smaller regions in the figures with significant coherence should not be over-interpreted as these occur in some 5 % of cases by chance.

3.3 Wavelet coherence significance testing of multiple models at a single site

Comparing significant wavelet coherence among US-Ha1 NEE and the output of multiple models (choosing SiBCASA, ED2, LoTEC, and ORCHIDEE, Fig. 3) reveals that the observed annual variability of NEE tends to be well-replicated

by the models. This finding is expected given the dominant role of orbital motions in controlling climate and flux in the temperate zone on these timescales. We note that Fig. 3 and subsequent figures ignore timescales smaller than one day to facilitate comparison between models that run on the daily and sub-daily time steps, and to emphasize longer timescales in the wavelet coherence significance tests.

Figure 3 also demonstrates that results from some models are significantly related to measurements at different times and timescales. LoTEC in particular is frequently related to observations on weekly and monthly timescales, but LoTEC is the only NACP model that implemented a data assimilation procedure, and should be expected to have a stronger relationship to measurements (Schwalm et al., 2010). Significant wavelet coherence exists among US-Ha1 NEE measurements and the SiBCASA, ED2 and LoTEC models, but not ORCHIDEE, on the seasonal timescale (one to several months) before 2002. Such findings question whether common model attributes (Table 2) are responsible for good fit or poor fit during these times and timescales.

A major advantage of converting the wavelet coherence values into binary significance maps is that the output of different models for the different measurement sites can be averaged or summed to explore aggregate model performance (e.g., via Eq. 2 or other metrics). We can begin by summing significant wavelet coherence for all 15 models that were run for US-Ha1 (Fig. 4, see Dietze et al., 2011). Figure 4 demonstrates that models are significantly related to NEE on the annual timescale for at least part of the measurement

Fig. 3. Same as Fig. 2 but showing significant wavelet coherence at timescales greater than 2 days between NEE measurements at Harvard Forest (US-Ha1) and the **(A)** SiBCASA, **(B)** ED2, **(C)** LoTEC and **(D)** ORCHIDEE model simulations.

period. More than 10 models are significantly related to the measurements on seasonal timescales, and frequent periods when multiple models are significantly related to measurements appear on weekly and monthly timescales. These features may be related to model structural attributes that can guide model testing and interpretation. We demonstrate such an approach by first exploring further the NEE observations and model output for US-Ha1. We then proceed to interpret results from the other nine research sites evaluated in this analysis (Table 1).
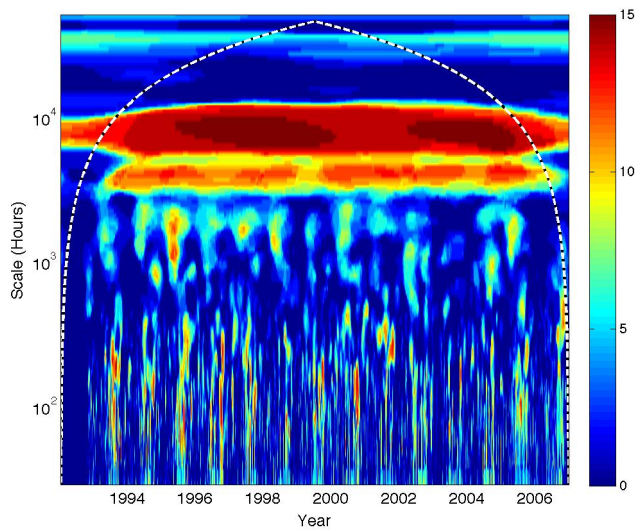
3.4 The role of common model features in determining significant wavelet coherence

Models in the NACP synthesis share features in common (Table 2, Schwalm et al., 2010). The role of these features in model performance across time and scale can be explored using the binary wavelet coherence significance approach demonstrated in Figs. 2 and 3. Logical model attributes to explore follow the findings of Schwalm et al. (2010) and include comparisons between prescribed versus prognostic

canopy phenology, the calculation of NEE as GPP minus ER or as NPP minus HR, the inclusion of foliar nitrogen in the model, and a comparison of models that use data assimilation (here namely LoTEC) versus those that do not (Table 2).

The results of the combined wavelet coherence significance analysis for US-Ha1 are shown in Fig. 5. In Fig. 5a, regions in time–timescale space for which coherence between NEE measurements from US-Ha1 and all models with prognostic phenology is significant, noting uncertainty, are given the value of 1 (see Eq. 2). From this, significant regions for which all models with prescribed phenology were subtracted. If all models with prognostic phenology are significantly related to NEE measurements for a given region in time and scale, and none with prescribed phenology are significant, the value of Eq. (2) equals 1−0 = 1 (dark blue). If the opposite holds, then the region equals negative 1 and is shown in dark red. This procedure is repeated for the different model attributes investigated (Table 2).

For example, from Fig. 5a, all (or most) models with prognostic phenology are often significantly related to NEE

Fig. 4. The sum of significant wavelet coherence coefficients for net ecosystem exchange (NEE) observations at the Harvard Forest (US-Ha1) and simulations by 15 ecosystem models (Table 2). This figure represents the sum of the subplots of Fig. 3 including the other models that were run at US-Ha1 listed in Dietze et al. (2011). A value of 15 indicates that all 15 models explored here are significantly related to NEE observations.

observations from US-Ha1 during seasonal timescales, especially earlier in the measurement period. These results suggest that phenology models work well in simulating the seasonal patterns that they seek to replicate, but prescribing phenology using remote sensing observations results in model patterns that are not significantly related to the pattern of measurements at these times and timescales. Schwalm et al. (2010) found that models with prognostic phenology and (to a lesser degree) those that calculate NEE as GPP minus ER tend to show better performance across sites (Fig. 5a and b). When this holds at US-Ha1, it is on timescales between ca. $10^{2.9}$ h (i.e., one month) and $10^{3.5}$ h (i.e., 4 months), which are the intermediate timescales on which model performance tends to diverge as identified by Dietze et al. (2011). This analysis reveals that the model attributes identified by Schwalm et al. (2010) as advantageous, prognostic phenology and NEE = GPP − ER, correspond to better performance on monthly to seasonal timescales at US-Ha1. Interestingly, including foliar N in the model did not result in unambiguous model improvement (Fig. 5c), and there were times and timescales when excluding N from the model resulted in improved model fit. The LoTEC model, with the data assimilation procedure, resulted in improved fit to data at many daily to seasonal timescales, but not annual or interannual timescales (Fig. 5d), noting that most models simulated annual variability correctly (Fig. 4). Interpreting significance across all sites and model attributes at all times and timescales is beyond the scope of this analysis, and we focus the remainder of our comparison on the dominant features

of the combined wavelet coherence significance analysis for different models and sites (Tables 1 and 2).
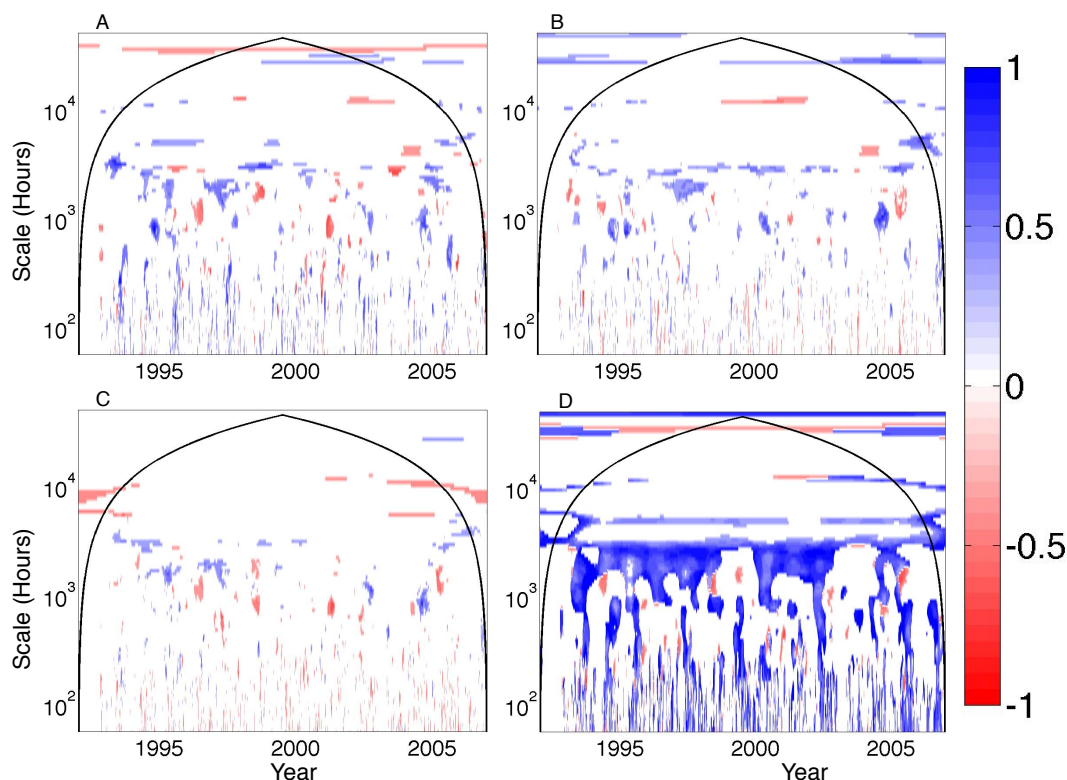
3.5 Phenology

Ecosystem models often fail to replicate the timing of spring green-up and autumn leaf senescence (Richardson et al., 2012), and, interestingly, incorporating satellite remote sensing data (i.e., prescribing phenology in models) may not represent an improvement in capturing phenological events (Fisher et al., 2007). However, NACP model results from the 10 study sites indicate that prescribing the phenology of leaf area index (LAI) often improves modeled carbon fluxes on seasonal and annual timescales at the cold, non-forest sites (i.e., CA-Let and CA-Mer, the deciduous forests CA-Oas and US-Ha1, and to a lesser degree US-UMB), and the agricultural ecosystem US-Ne1 (Figs. 5 and 6). Models that include prognostic phenology are often significantly related to NEE observations at seasonal and annual timescales at the coniferous ecosystems in the temperate continental climate zone: CA-Obs and US-Ho1.

Despite many successes, prescribing phenology resulted in erroneous model fit in some ecosystems, times, and timescales (Fig. 6), in agreement with Richardson et al. (2012), who found that model biases of two weeks or more were common for deciduous forests. Predicting phenology in the coniferous forests (CA-Obs and US-Ho1) is a superior strategy for modeling NEE on seasonal timescales. This makes sense given the difficulty of using remote sensing to detect seasonal changes in leaf area and photosynthetic activity in evergreen canopies. The creation of effective prognostic phenology models for grasslands and croplands remains challenging, especially when cropping systems often depend on the decisions of land managers. Remote sensing is often unsuccessful for capturing grassland phenology (Reed et al., 1994), due in part to the fact that the shift from green to brown biomass is critical for modeling NEE but can be subtle and difficult to ascertain remotely (Sus et al., 2010), and, regardless, photosynthetic activity may decline long before leaf spectral indices begin to change (Bauerle et al., 2012).

3.6 NEE calculation

Models calculate ecosystem carbon uptake and loss in different ways, and the NACP models can roughly be categorized as those that calculate NEE as GPP minus ER and those that calculate NEE as NPP minus HR (Schwalm et al., 2010). Models that calculate NEE as NPP minus HR tend to fit better than models that calculate NEE as GPP minus ER on the annual timescale at the Canadian grassland (CA-Let) and bog (CA-Mer) sites, which are characterized by short-statured vegetation and pronounced seasonality (Fig. 7). Models that calculate NEE as NPP minus HR also represent an improvement on seasonal and annual timescales at the deciduous forests Ca-Oas and US-UMB, and at daily
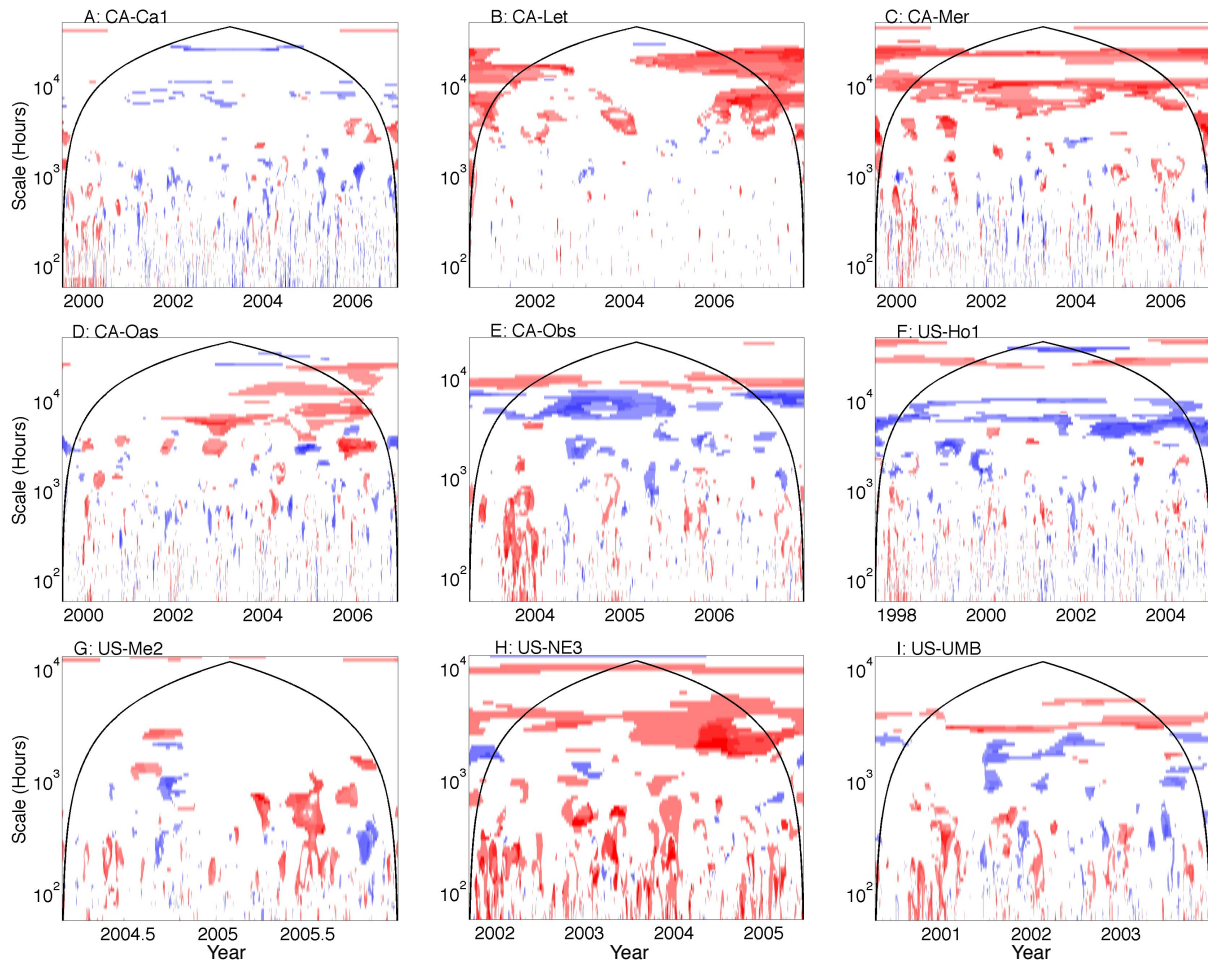
Fig. 5. The ratio of significant wavelet coherence for different model attributes for the net ecosystem exchange (NEE) observations at the Harvard Forest (US-Ha1) following Eq. (1). Areas of dark blue represent times and scales where all models that include prognostic phenology **(A)**, the NEE calculation as GPP minus ER **(B)**, the inclusion of nitrogen **(C)**, and a model that used data assimilation (LoTEC, **D**) are significantly related to NEE observations, and when none of the opposing model strategy listed in Table 2 is significant. Areas of dark red represent periods when the opposite holds.

to weekly timescales at the coniferous forests Ca-Obs, US-Ho1, and US-Me2. Models that calculate NEE as GPP minus ER tend to fit better on monthly to seasonal timescales at the coniferous forests CA-Obs and US-Ho1. In general, simulating NEE and HR results in poorer NEE model fit at seasonal and annual timescales in coniferous stands, and simulating GPP and ER presents more of a challenge in grasslands, wetlands, and deciduous forests (Vargas et al., 2013). Many of the subplots in Fig. 7 show a scale-wise shift (from red to blue or vice versa) as one moves to longer scales in time, suggesting that the responses of GPP, NPP, ER and HR to environmental drivers that act on different timescales need to be examined carefully for proper frequency response.

3.7 Nitrogen

Models utilizing measurements of foliar N show improved fits on interannual timescales compared models that exclude N at a coniferous forest (US-Ho1; Fig. 8f) and to a lesser degree at a deciduous forest (CA-Oas, Fig. 8d). This finding supports the incorporation of canopy N as an important component for accurately modeling spatial and temporal patterns in NEE (Hollinger et al., 2009; Ollinger and Smith, 2005;

Ollinger et al., 2008). However, it is discouraging that incorporating N improves interannual model fit for only a couple of sites rather than for all sites; note for example the poor fit of models that include N on timescales shorter than the interannual timescale at Ca-Oas (Fig. 8d). Climatic variables tend to be unrelated or weakly related to observed NEE on interannual timescales (Stoy et al., 2009), and variability in biological drivers like canopy N is thought to be a principle control over NEE variability on interannual timescales (Richardson et al., 2007). The role of biological lags (e.g., growth and NPP lagging behind C uptake) tends to be poorly represented in the current generation of ecosystem models (Keenan et al., 2012), as are the dynamics of the non-structural carbohydrates that can contribute to such lags (Gough et al., 2009, 2010; Richardson et al., 2013). Modeling the biological responses to interannual climatic variability continues to be a major research challenge (Richardson et al., 2007; Siqueira et al., 2006), and it appears that modeling N improves models of NEE, but only in certain instances. Including foliar N improves model fit on certain timescales for different sites; for example including N appears to improve models in CA-Let, CA-Oas and CA-Obs, during summer months in 2006. The summer of 2006 was at the time the second warmest on
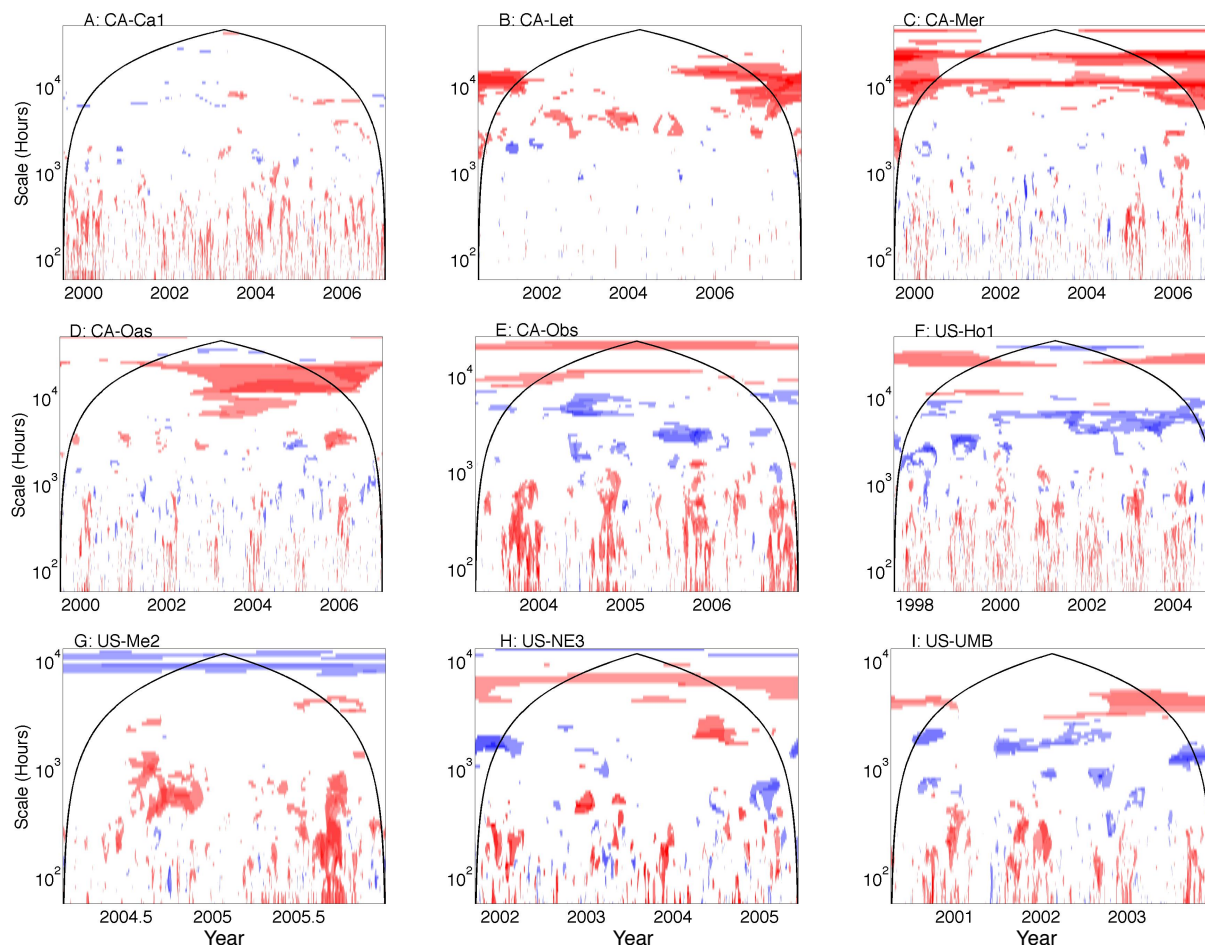
Fig. 6. The ratio of wavelet coherence significance tests for models with prognostic prescribed phenology for nine sites in the North American Carbon Program Interim Synthesis. The color bar follows Fig. 5. Areas of dark blue represent times and scales for which models that use prognostic leaf area index (LAI) are significantly related to NEE measurements and those that use prescribed LAI are not significantly related to NEE measurements. Regions for which models that use prescribed LAI are significant and those that use prognostic LAI are not significant are shown as dark red. The color bar follows Fig. 5.

record in Canada, but the role of N in improving modeled NEE during these conditions is difficult to interpret and may be due to correct hydrological response regardless of the N subroutine in the models.

3.8 Data assimilation

Data assimilation for formally fusing observations and models has gained increased attention in the biogeosciences (Dietze et al., 2013; Hill et al., 2011; Rastetter et al., 2010; Raupach et al., 2005; Williams et al., 2005). LoTEC applied a data assimilation procedure in the NACP modeling exercise, and output in many instances represented a striking improvement against the aggregate output of other models (Fig. 9). Namely, LoTEC output is significantly related to NEE measurements across many timescales at the deciduous forests CA-Oas and US-UMB. LoTEC also demon-

strated improved fit compared to other models at seasonal and/or annual timescales at the coniferous forests CA-Obs and US-Ho1 and the crop US-Ne1. LoTEC was not significantly related to NEE measurements (and the average of other models were) across most sites with the exception of CA-Oas and US-UMB. Results suggest that the optimized parameters computed in the LoTEC data assimilation procedure can improve fit across times and timescales, especially for some of the ecosystems that exhibit pronounced seasonality in canopy dynamics (i.e., some deciduous forests, and the agricultural ecosystem). Results also demonstrate that data assimilation does not always result in significant relationships between measurements and models; there are many periods, often timescales between a day and about a month and a half (10^3 h) and at interannual timescales, where LoTEC is not significantly related to measurements. Such findings demonstrate the importance of model parameterization, and

Fig. 7. The ratio of wavelet coherence significance tests for models with different calculations of the net ecosystem exchange (NEE). The color bar follows Fig. 5. Areas of dark blue represent times and scales for which models that calculate NEE as gross primary productivity (GPP) minus ecosystem respiration (ER) are significantly related to NEE measurements and models that calculate NEE as net primary productivity (NPP) minus heterotrophic respiration (HR) are not significantly related to NEE measurements. Regions for which models that calculate NEE as NPP minus HR are significant and GPP minus ER are not significant are dark red. The color bar follows Fig. 5.
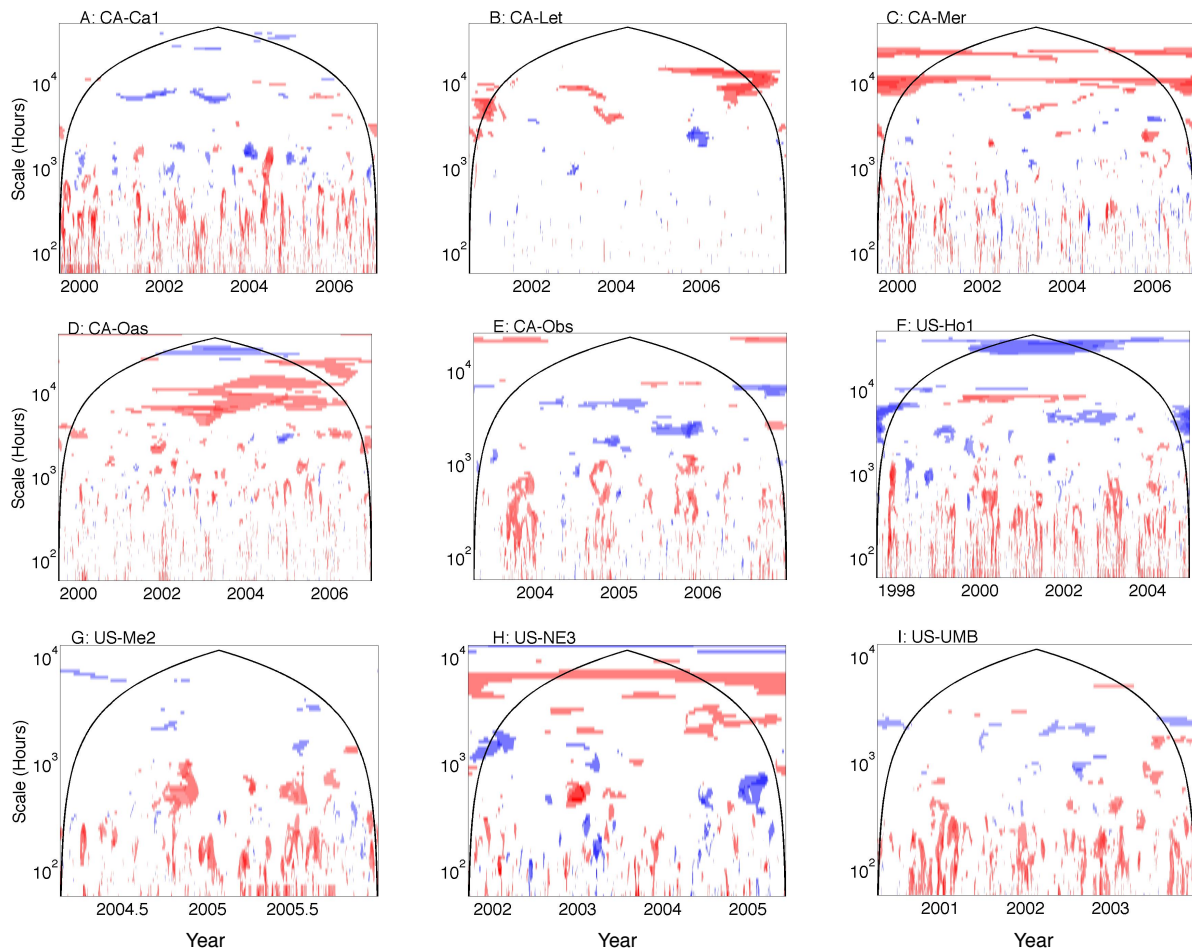
also suggest that data assimilation should not take the place of efforts to improve model structure.

3.9 The analysis of models at multiple timescales

We used wavelet coherence as a criterion for model/measurement comparison in this study. Spectral analyses can also be used to discriminate among model subroutines and inputs (Stoy et al., 2005) or demonstrate model improvement (Williams et al., 2009), and it is for these purposes that wavelet coherence may find the most application in the biogeosciences. Wang et al. (2011) recently used wavelet coherence to quantify patterns of CABLE model output (Kowalczyk et al., 2006) and demonstrated how model refinements improved predictions of NEE, latent heat and sensible heat on multiple timescales, although observed patterns in interannual variability in NEE remained difficult for CABLE to resolve. Wavelet coherence has also

been applied to identify model/measurement mismatches resulting from phenological response to monsoon activity (Hong and Kim, 2011) and for detailed studies on the controls over carbon cycling in individual sites (van Gorsel et al., 2013). We suggest that any comprehensive model diagnostic toolkit should explore model frequency response, and we demonstrate the application of wavelet coherence as a model–measurement comparison technique that is also visually intuitive. It is important to note that a wavelet coherence test for matches in patterns, rather than magnitudes, and by itself is an incomplete metric for model fidelity. The potential for wavelet coherence to identify regions of false positive coherence erroneously must be acknowledged when interpreting results (Maraun et al., 2007).

Future research efforts should compare wavelet-based approaches with other time-series decomposition techniques, including singular systems analysis (Mahecha et al., 2010),
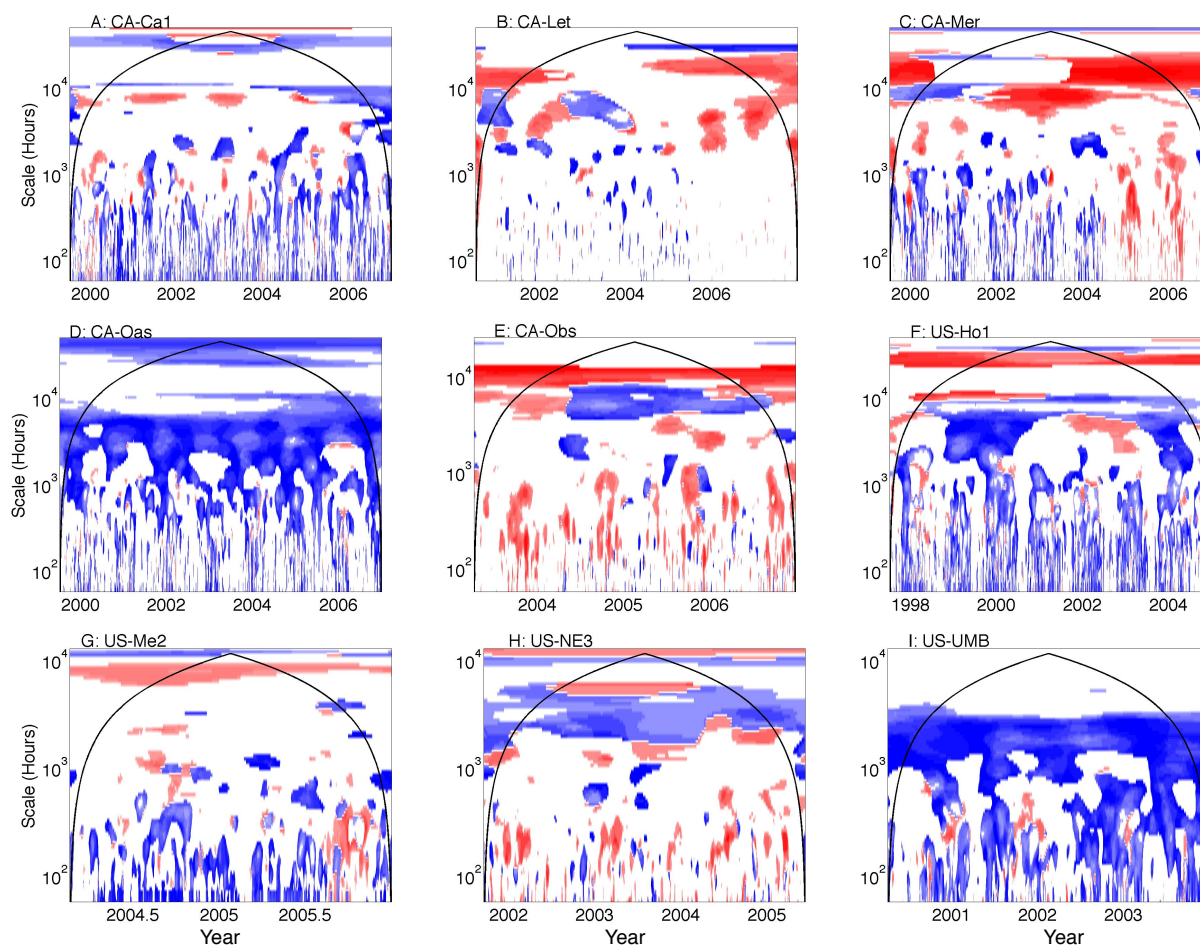
Fig. 8. The ratio of wavelet coherence significance tests for models that include or exclude foliar nitrogen (N). The color bar follows Fig. 5. Areas of dark blue represent times and scales for which models that incorporate N are significantly related to NEE measurements and models that exclude N are not significantly related to NEE measurements. Regions for which models that exclude N are significantly related to NEE measurements and those that include N are not are dark red. The color bar follows Fig. 5.

spectral analysis of model residuals (Dietze et al., 2011; Vargas et al., 2010, 2013), and/or quantification of causal relationships among measurements and models across time and spectra using the Granger definition (Detto et al., 2012). Here, it is pertinent to point out that different approaches to creating null spectra or surrogate time series for significance testing are favored for different time-series methodologies, and to reiterate that the estimation of significance resides partially with the choice of null or surrogate spectra. Recent studies in the biogeosciences provide examples; Molini et al. (2010) and Detto et al. (2012) applied iterative amplitude adjusted Fourier transform (IAAFT) surrogates (Schreiber and Schmitz, 1996, 2000; Venema et al., 2006) to create signals that preserve the probability distribution, power spectra, and linear correlation structure of the measurements for determining significant spectral Granger causality in rainfall and canopy assimilation and soil respiration time series. Here, such an approach would shuffle the deterministic periodic-

ities at the diurnal and seasonal timescales and result in impossible time series against which to ascertain significance; hence the original red noise spectra selected by Grinsted et al. (2004) were also chosen.

4 Conclusions

We demonstrated an application of wavelet coherence for testing significant relationships between flux observation and the output of multiple ecosystem models run at multiple different study sites. Models with prognostic phenology were often significantly related to NEE measurements on seasonal timescales in coniferous sites, but models with prescribed phenology improved seasonal and annual model fit in grassland and wetland study sites, and to a lesser degree in the deciduous forests US-Ha1 and US-UMB. The inclusion of foliar N improved model performance on interannual timescales at US-Ho1.

Fig. 9. The ratio of wavelet coherence significance tests for the model using a data assimilation procedure (LoTEC) versus other models that do not use data assimilation. The color bar follows Fig. 5. Periods and scales in time for which LoTEC is significantly related to NEE measurements and other models are not significantly related to NEE measurements (on average) are shown as dark blue. Regions for which models other than LoTEC are significantly related to NEE measurements and LoTEC is not significantly related to NEE measurements are shown as dark red. The color bar follows Fig. 5.

Model pattern tended to match observed NEE on diurnal timescales during the growing season and on annual timescales (e.g., Figs. 1 and 2), but previous analyses indicate that models often misrepresent the magnitude of fluxes on these highly energetic timescales (Dietze et al., 2011). Despite correct frequency responses on growing-season diurnal and annual timescales as we find here, Dietze et al. (2011) demonstrated that proper parameterization of flux magnitude on these scales should remain a focus of modeling efforts. LoTEC results (Figs. 3 and 9) hint that data assimilation can improve model fit on the intermediate weekly to seasonal timescales during many periods, but modeled flux variability on diurnal and interannual timescales was often not significantly related to measurements, suggesting that models still require mechanistic improvements even under parameter sets optimized by the data assimilation routine. Mechanistic explanations for describing interannual NEE variability still elude most models, although correctly modeling N dynamics may be a strategy for progressing on this problem in some ecosystems (Fig. 8).

Wavelet coherence adds an additional diagnostic tool to a modeler's conceptual toolbox for evaluating the performance of single models or suites of models (Grinsted et al., 2004; Torrence and Compo, 1998; Williams et al., 2009). Future efforts should determine the benefits and drawbacks of wavelet, Fourier, and singular systems analysis approaches for model/measurement comparisons (Katul et al., 2001; Mahecha et al., 2010; Siqueira et al., 2006; Maraun et al., 2007; Vargas et al., 2010), use the outcomes of multiple spectral analyses to provide insight into how and why models fail, and use this information to improve model performance at the multiple times and timescales at which biogeochemical fluxes vary.

References

Akaike, H.: A new look at the statistical model identification, IEEE T. Automat. Contr., 19, 716–723, 1974.

Arain, M. A., Yaun, F., and Black, T. A.: Soil-plant nitrogen cycling modulated carbon exchanges in a western temperate conifer forest in Canada, Agr. Forest Meteorol., 140, 171–192, 2006.

Baker, I. T., Prihodko, L., Denning, A. S., Goulden, M., Miller, S., and da Rocha, H. R.: Seasonal drought stress in the Amazon: Reconciling models and observations, J. Geophys. Res., 113, G00B01, doi:10.1029/2007JG000644, 2008.

Barr, A. G., Hollinger, D. Y., Richardson, A. D.: CO_2 flux measurement uncertainty estimates for NACP, Eos Trans. AGU, 90, Fall Meet. Suppl., Abstract B54A-04.

Bauerle, W. L., Oren, R., Way, D. A, Qian, S. S., Stoy, P. C., Thornton, P. E., Bowden, J. D., Hoffman, F. M., and Reynolds, R. F.: Photoperiodic regulation of the seasonal pattern of photosynthetic capacity and the implications for carbon cycling, P. Natl. Acad. Sci. USA, 109, 8612–8617. doi:10.1073/pnas.1119131109, 2012.

Causarano, H. J., Shaw, J. M., Franzluebbers, A. J., Reeves, D. W., Raper, R. L., Balkcom, K. S., Norfleet, M. L., and Izaurralde, R. C.: Simulating field-scale soil organic carbon dynamics using EPIC, Soil Sci. Soc. Am. J., 71, 1174–1185, 2007.

Detto, M., Molini, A., Katul, G. G., Stoy, P. C., Palmroth, S., and Baldocchi, D. D.: Causality and persistence in ecological systems: a nonparametric spectral Granger causality approach, Am. Nat., 179, 524–535, 2012.

Dietze, M. C, Vargas, R., Richardson, A. D., Stoy, P. C., Barr, A. G., Anderson, R. S., Arain, A., Baker, I. T., Black, T. A., Chen, J. M., Ciais, P., Flanagan, L. B., Gough, C. M., Grant, R. F., Hollinger, D. Y., Izaurralde, C., Kucharik, C. J., Lafleur, P. M., Liu, S., Lokupitiya, E., Luo, Y., Munger, J. W., Peng, C., Poulter, B., Price, D. T., Ricciuto, D. M., Riley, W. J., Sahoo, A. K., Schaefer, K., Tian, H., Verbeeck, H., and Verma, S.

B.: Characterizing the performance of ecosystem models across time scales: A spectral analysis of the North American Carbon Program site-level synthesis, J. Geophys. Res., 116, G04029, doi:10.1029/2011JG001661, 2011.

Dietze, M. C., LeBauer, D. S., and Kooper, R.: On improving the communication between models and data, Plant Cell Environ., 36, 1575–1585, doi:10.1111/pce.12043, 2013.

Fisher, J. I., Richardson, A. D., and Mustard, J. F.: Phenology model from surface meteorology does not capture satellite-based greenup estimations, Glob. Change Biol., 13, 707–721, doi:10.1111/j.1365-2486.2006.01311.x, 2007.

Flanagan, L. B., Wever, L. A., and Carlson, P. J.: Seasonal and inter-annual variation in carbon dioxide exchange and carbon balance in a northern temperate grassland, Glob. Change Biol., 8, 599–615, 2002.

Friedlingstein, P., Cox, P., Betts, R. A., Bopp, L., von Blow, W., Brovkin, V., Cadule, P., Doney, S. C., Eby, M., Fung, I. Y., Bala, G., John, J., Jones, C. D., Joos, F., Kato, T., Kawamiya, M., Knorr, W., Lindsay, K., Matthews, H. D., Raddatz, T., Rayner, P. J., Reick, C., Roeckner, E., Schnitzler, K.-G., Schnur, R., Strassmann, K., Weaver, A. J., Yoshikawa, C., and Zeng, N.: Climate-carbon cycle feedback analysis: results from the C4MIP model intercomparison, J. Climate, 19, 3337–3353, 2006.

Gough, C. M., Vogel, C. S., Schmid, H. P., Su, H.-B., and Curtis, P. S.: Multi-year convergence of biometric and meteorological estimates of forest carbon storage, Agr. Forest Meteorol., 148, 158–170, 2008.

Gough, C. M., Flower, C. E., Vogel, C. S., Dragoni, D., and Curtis, P. S.: Whole-ecosystem labile carbon production in a north temperate deciduous forest, Agr. Forest Meteorol., 149, 1531–1540, 2009.

Gough, C. M., Flower, C. E., Vogel, C. S., and Curtis, P. S.: Phenological and temperature controls on the temporal non-structural carbohydrate dynamics of *Populus grandidentata* and *Quercus rubra*, Forests, 1, 65–81, 2010.

Grant, R. F., Arain, A., Arora, V., Barr, A., Black, T. A., Chen, J., Wang, S., Yuan, F., and Zhang, Y.: Intercomparison of techniques to model high temperature effects on CO_2 and energy exchange in temperate and boreal coniferous forests, Ecol. Model., 188, 217–252, 2005.

Gray, J.: Jim Gray on eScience: A transformed scientific method, in: The fourth paradigm: Data-intensive scientific discovery, edited by: Hey, T., Tansley, S., and Tolle, K., Microsoft Research, 284, 2009.

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

Grinsted, A., Moore, J. C., and Jevrejeva, S.: Application of the cross wavelet transform and wavelet coherence to geophysical time series, Nonlinear Proc. Geoph., 11, 561–566, 2004.

Hanson, P. J., Amthor, J. S., Wullschleger, S. D., Wilson, K. F., Grant, R. F., Hartley, A., Hui, D. F., Hunt, E. R. J., Johnson, D. W., Kimball, J. S., King, A. W., Y., L., McNulty, S. G., Sun, G., Thornton, P. E., S., W., Williams, M., Baldocchi, D. D., and Cushman, R. M.: Oak forest carbon and water simulations: model intercomparisons and evaluations against independent data., Ecol. Monogr., 74, 443–489, 2004.

Hill, T. C., Quaife, T., and Williams, M.: A data assimilation method for using low-resolution Earth observation data in heterogeneous ecosystems, J. Geophys. Res., 116, D08117, doi:10.1029/2010jd015268, 2011.

Hollinger, D. Y., Ollinger, S. V., Richardson, A. D., Meyers, T., Dail, D. B., Martin, M. E., Scott, N. A., Arkebauer, T. J., Baldocchi, D. D., Clark, K. L., Curtis, P. S., Davis, K. J., Desai, A. R., Dragoni, D., Goulden, M. L., Gu, L., Katul, G. G., Pallardy, S. G., Paw U, K. T., Schmid, H. P., Stoy, P. C., Suyker, A. E., and Verma, S. B.: Albedo estimates for land surface models and support for a new paradigm based on foliage nitrogen concentration., Glob. Change Biol., 16, 696–710, 2009.

Hong, J. and Kim, J.: Impact of the Asian monsoon climate on ecosystem carbon and water exchanges: a wavelet analysis and its ecosystem modeling implications, Glob. Change Biol., 17, 1900–1916, 2011.

Katul, G. G., Lai, C.-T., Schäfer, K. V. R., Vidakovic, B., Albertson, J. D., Ellsworth, D. S., and Oren, R.: Multiscale analysis of vegetation surface fluxes: from seconds to years, Adv. Water Resour., 24, 1119–1132, 2001.

Keenan, T. F., Baker, I., Barr, A., Ciais, P., Davis, K., Dietze, M., Dragoni, D., Gough, C. M., Grant, R., Hollinger, D., Hufkens, K., Poulter, B., McCaughey, H., Raczka, B., Ryu, Y., Schaefer, K., Tian, H., Verbeeck, H., Zhao, M., and Richardson, A. D.: Terrestrial biosphere model performance for inter-annual variability of land-atmosphere CO_2 exchange, Glob. Change Biol., 18, 1971–1987, 2012.

Kowalczyk, E. A., Wang, Y. P., Law, R. M., Davies, H. L., McGregor, J. L., and Abramowitz, G.: The CSIRO atmosphere biosphere land exchange (CABLE) model for use in climate models and as an offline model, CSIRO Marine and Atmospheric Research, ISBN 1 921232 39 0, 2006.

Krinner, G., Viovy, N., de Noblet-Ducoudré, N., Ogée, J., Polcher, J., Friedlingstein, P., Ciais, P., Sitch, S., and Prentice, I. C.: A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system, Global Biogeochem. Cy., 19, GB1015, doi:10.1029/2003GB002199, 2005.

Krishnan, P., Black, T. A., Grant, N. J., Barr, A. G., Hogg, E. H., Jassal, R. S., Morgenstern, K.: Impact of changing soil moisture distribution on net ecosystem productivity of a boreal aspen forest during and following drought, Agr. For. Meteorol. 139, 208–223, 2006.

Kucharik, C. J. and Twine, T. E.: Residue, respiration and residuals: Evaluation of a dynamic agroecosystem model using eddy flux measurements and biometric data, Agr. Forest Meteorol., 146, 134–158, 2007.

Lafleur, P. M., Roulet, N. T., Bubier, J. L., Moore, T. R., and Frolking, S.: Interannual variability in the peatland-atmosphere carbon dioxide exchange at an ombrotrophic bog, Global Biogeochem. Cy., 17, 1036, doi:10.1029/2002GB001983, 2003.

Law, B. E., Arkebauer, T. J., Campbell, J. L., Chen, J., Sun, O., Schwartz, M., van Ingen, C., and Verma, S.: Terrestrial carbon observations: Protocols for vegetation sampling and data submission, Report 55, FAO, Rome, 87, 2008.

Li, H., Qiu, J., Wang, L., Tang, H., Li, C., and Van Ranst, E.: Modelling impacts of alternative farming management practices on greenhouse gas emissions from a winter wheat-maize rotation system in China, Agr. Ecosyst. Environ., 135, 24–33, 2010.

Light, A. and Bartlein, P. J.: The end of the rainbow? Color schemes for improved data graphics, Eos Trans. AGU, 85, 385–391, 2004.

Liu, J., Chen, J. M., Cihlar, J., and Chen, W.: Net primary productivity distribution in the BOREAS region from a process model using satellite and surface data, J. Geophys. Res., 104, 27735–27754, 1999.

Lokupitiya, E., Denning, S., Paustian, K., Baker, I., Schaefer, K., Verma, S., Meyers, T., Bernacchi, C. J., Suyker, A., and Fischer, M.: Incorporation of crop phenology in Simple Biosphere Model (SiBcrop) to improve land-atmosphere carbon exchanges from croplands, Biogeosciences, 6, 969–986, doi:10.5194/bg-6-969-2009, 2009.

Mahecha, M. D., Reichstein, M., Jung, M., Senevirante, S. I., Zaehle, S., Beer, C., Braakhekke, M. C., Carvalhais, N., Lange, H., Le Maire, G., and Moors, E.: Comparing observations and process-based simulations of biosphere-atmosphere exchanges on multiple timescales, J. Geophys. Res., 115, G02003, doi:10.1029/2009JG001016, 2010.

Maraun, D. and Kurths, J.: Cross wavelet analysis: significance testing and pitfalls, Nonlinear Proc. Geoph., 11, 505–514, 2004.

Maraun, D., Kurths, J., and Holschneider, M.: Nonstationary Gaussian processes in wavelet domain: Synthesis, estimation, and significance testing, Physical Review E, 75, 16707, 2007.

Medvigy, D., Wofsy, S. C., Munger, J. W., Hollinger, D. Y., and Moorcroft, P. R.: Mechanistic scaling of ecosystem function and dynamics in space and time: Ecosystem Demography model version 2, J. Geophys. Res., 114, G01002, doi:10.1029/2008JG000812, 2009.

Molini, A., Katul, G. G., and Porporato, A.: Causality across rainfall time scales revealed by continuous wavelet transforms, Journal Geophys. Res., 115, D14123, doi:10.1029/2009JD013016, 2010.

Ollinger, S. V. and Smith, M.-L.: Net primary production and canopy nitrogen in a temperate forest landscape: an analysis using imaging spectroscopy, modeling and field data, Ecosystems, 8, 760–778, 2005.

Ollinger, S. V., Richardson, A. D., Martin, M. E., Hollinger, D. Y., Frolking, S., Reich, P. B., Plourde, L. C., Katul, G. G., Munger, J. W., Oren, R., Smith, M.-L., Paw U, K. T., Bolstad, P. V., Cook, B. D., Day, M. C., Martin, T. A., Monson, R. K., and Schmid, H. P.: Canopy nitrogen, carbon assimilation, and albedo in temperate and boreal forests: functional relations and potential climate feedbacks, P. Natl. Acad. Sci. USA, 105, 19336–19341, doi:10.1073/pnas.0810021105, 2008.

Papale, D., Reichstein, M., Aubinet, M., Canfora, E., Bernhofer, C., Kutsch, W., Longdoz, B., Rambal, S., Valentini, R., Vesala, T., and Yakir, D.: Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: algorithms and uncertainty estimation, Biogeosciences, 3, 571–583, doi:10.5194/bg-3-571-2006, 2006.

Rastetter, E. B., Williams, M., Griffin, K. L., Kwiatkowski, B. L., Tomasky, G., Potosnak, M. J., Stoy, P. C., Shaver, G. R., Stieglitz, M., Hobbie, J. E., and Kling, G. W.: Processing arctic eddy-flux data using a simple carbon-exchange model embedded in the ensemble Kalman filter, Ecol. Appl., 20, 1285–1301, doi:10.1890/09-0876.1, 2010.

Raupach, M. R., Rayner, P. J., Barrett, D. J., DeFries, R. S., Heimann, M., Ojima, D., Quegan, S., and Schmullius, C. C.: Model-data synthesis in terrestrial carbon observation: meth-

ods, data requirements and data uncertainty specifications, Glob. Change Biol., 11, 378–397, 2005.

Reed, B. C., Brown, J. F., VanderZee, D., Loveland, T. R., Merchant, J. W., and Ohlen, D. O.: Measuring phenological variability from satellite imagery, J. Veg. Sci., 5, 703–714, doi:10.2307/3235884, 1994.

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T. G., Granier, A., Grünwald, T., Havránková, 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., Yakier, D., and Valentini, R.: On the separation of net ecosystem exchange into assimilation and ecosystem respiration: Review and improved algorithm, Glob. Change Biol., 11, 1424–1439, 2005.

Ricciuto, D. M., Thornton, P. E., Schaefer, K., Cook, R. B., and Davis, K. J.: How uncertainty in gap-filled meteorological input forcing at eddy covariance towers impacts modeled carbon and energy flux, Eos Trans. AGU, 90, Fall Meet. Suppl., Abstract B54A-03, 2009.

Richardson, A. D., Hollinger, D. Y., Aber, J. D., Ollinger, S. V., and Braswell, B. H.: Environmental variation is directly responsible for short- but not long-term variation in forest-atmosphere carbon exchange, Glob. Change Biol., 13, 788–803, 2007.

Richardson, A. D., Mahecha, M. D., Falge, E., Kattge, J., Moffat, A. M., Papale, D., Reichstein, M., Stauch, V. J., Braswell, B. H., Churkina, G., Kruijt, B., and Hollinger, D. Y.: Statistical properties of random CO_2 flux measurement uncertainty inferred from model residuals, Agr. Forest Meteorol., 148, 38–50, 2008.

Richardson, A. D., Hollinger, D. Y., Dail, D. B., Lee, J. T., Munger, J. W., and O'Keefe, J.: Influence of spring phenology on seasonal and annual carbon balance in two contrasting New England forests, Tree Physiol., 29, 321–331, 2009.

Richardson, A. D., Anderson, R. S., Arain, M. A., Barr, A. G., Bohrer, G., Chen, G., Chen, J. M., Ciais, P., Davis, K. J., Desai, A. R., Dietze, M. C., Dragoni, D., Garrity, S. R., Gough, C. M., Grant, R., Hollinger, D. Y., Margolis, H. A., McCaughey, H., Migliavacca, M., Monson, R. K., Munger, J. W., Poulter, B., Raczka, B. M., Ricciuto, D. M., Sahoo, A. K., Schaefer, K., Tian, H., Vargas, R., Verbeeck, H., Xiao, J., and Xue, Y.: Terrestrial biosphere models need better representation of vegetation phenology: results from the North American Carbon Program Site Synthesis, Glob. Change Biol., 18, 566–584, doi:10.1111/j.1365-2486.2011.02562.x, 2012.

Richardson, A. D., Carbone, M. S., Keenan, T. F., Czimczik, C. I., Hollinger, D. Y., Murakami, P., Schaberg, P. G., and Xu, X.: Seasonal dynamics and age of stemwood nonstructural carbohydrates in temperate forest trees, New Phytol., 197, 850–861, doi:10.1111/nph.12042, 2013.

Riley, W. J., Still, C. J., Torn, M. S., and Berry, J. A.: A mechanistic model of $H_2^{18}O$ and $C^{18}OO$ fluxes between ecosystems and the atmosphere: Model description and sensitivity analyses, Global Biogeochem. Cy., 16, 1095, doi:10.1029/2002GB001878, 2002.

Schaefer, K., Zhang, T., Slater, A. G., Lu, L., Etringer, A., and Baker, I.: Improving simulated soil temperatures and soil freeze/thaw at high-latitude regions in the Simple Biosphere/Carnegie-Ames-Stanford Approach model, J. Geophys. Res., 114, F02021, doi:10.1029/2008JF001125, 2009.

Schaefer, K., Schwalm, C. R., Williams, C., Arain, M. A., Barr, A., Chen, J. M., Davis, K. J., Dimitrov, D., Hilton, T. W., Hollinger, D. Y., Humphreys, E., Poulter, B., Raczka, B. M., Richardson, A. D., Sahoo, A., Thornton, P., Vargas, R., Verbeeck, H., Anderson, R., Baker, I., Black, T. A., Bolstad, P., Chen, J., Curtis, P. S., Desai, A. R., Dietze, M., Dragoni, D., Gough, C., Grant, R. F., Gu, L., Jain, A., Kucharik, C., Law, B., Liu, S., Lokipitiya, E., Margolis, H. A., Matamala, R., McCaughey, J. H., Monson, R., Munger, J. W., Oechel, W., Peng, C., Price, D. T., Ricciuto, D., Riley, W. J., Roulet, N., Tian, H., Tonitto, C., Torn, M., Weng, E., and Zhou, X.: A model-data comparison of gross primary productivity: Results from the North American Carbon Program site synthesis, J. Geophys. Res., 117, G03010, doi:10.1029/2012jg001960, 2012.

Schmid, H. P., Su, H. B., Vogel, C. S., and Curtis, P. S.: Ecosystem-atmosphere exchange of carbon dioxide over a mixed hardwood forest in northern lower Michigan, J. Geophys. Res.-Atmos., 108, 4417, doi:10.1029/2002JD003011, 2003.

Schreiber, T. and Schmitz, A.: Improved surrogate data for nonlinearity tests, Phys. Rev. Lett., 77, 635–638, 1996.

Schreiber, T. and Schmitz, A.: Surrogate time series, Physica D, 142, 346–382, 2000.

Schwalm, C. R., Black, T. A., Morgenstern, K., and Humphreys, E. R.: A method for deriving net primary productivity and component respiratory fluxes from tower-based eddy covariance data: a case study using a 17-year data record from a Douglas-fir chronosequence, Glob. Change Biol., 13, 370–385, 2007.

Schwalm, C. R., Williams, C. A., Schaefer, K., Anderson, R., Arain, M. A., Baker, I., Barr, A. G., Black, T. A., Chen, G., Chen, J. M., Ciais, P., Davis, K. J., Desai, A. R., Dietze, M., Dragoni, D., Fischer, M. L., Flanagan, L. B., Grant, R., Gu, L., Hollinger, D., Izaurralde, R. C., Kucharik, C. J., Lafleur, P. M., Law, B. E., Li, L., Li, Z., Liu, S., Lokipitiya, E., Luo, Y., Ma, S., Margolis, H., Matamala, R., McCaughey, J. H., Monson, R. K., Oechel, W., Peng, C., Poulter, B., Price, D. T., Riciutto, D. M., Riley, W., Sahoo, A. K., Sprintsin, M., Sun, J., Tian, H., Tonitto, C., Verbeeck, H., and Verma, S. B.: A model-data intercomparison of CO_2 exchange across North America: Results from the North American Carbon Program Site Synthesis, J. Geophys. Res, 115, G00H05, doi:10.1029/2009JG001229, 2010.

Siqueira, M. B. S., Katul, G. G., Sampson, D. A., Stoy, P. C., Juang, J.-Y., McCarthy, H. R., and Oren, R.: Multi-scale model inter-comparisons of CO_2 and H_2O exchange rates in a maturing southeastern U.S. pine forest, Glob. Change Biol., 12, 1189–1207, 2006.

Sitch, S., Smith, B., Prentice, I. C., Arneth, A., Bondeau, A., Cramer, W., Kaplan, J. O., Levis, S., Lucht, W., Sykes, M., Thonicke, K., and Venevsky, S.: Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model, Glob. Change Biol., 9, 161–185, 2003.

Stoy, P. C., Katul, G. G., Siqueira, M. B. S., Juang, J.-Y., McCarthy, H. R., Kim, H.-S., Oishi, A. C., and Oren, R.: Variability in net ecosystem exchange from hourly to inter-annual time scales at adjacent pine and hardwood forests: a wavelet analysis, Tree Physiol., 25, 887–902, 2005.

Stoy, P. C., Katul, G. G., Siqueira, M. B. S., Juang, J.-Y., Novick, K. A., Uebelherr, J. M. and Oren, R.: An evaluation of models for partitioning eddy covariance-measured net ecosystem exchange

into photosynthesis and respiration, Agr. Forest Meteorol., 141, 2–18, doi:10.1016/j.agrformet.2006.09.001, 2006.

Stoy, P. C., Richardson, A. D., Baldocchi, D. D., Katul, G. G., Stanovick, J., Mahecha, M. D., Reichstein, M., Detto, M., Law, B. E., Wohlfahrt, G., Arriga, N., Campos, J., McCaughey, J. H., Montagnani, L., Paw U, K. T., Sevanto, S., and Williams, M.: Biosphere-atmosphere exchange of CO_2 in relation to climate: a cross-biome analysis across multiple time scales, Biogeosciences, 6, 2297–2312, doi:10.5194/bg-6-2297-2009, 2009.

Sus, O., Williams, M., Bernhofer, C., Béziat, P., Buchmann, N., Ceschia, E., Doherty, R., Eugster, W., Grünwald, T., Kutsch, W., Smith, P., and Wattenbach, M.: A linked carbon cycle and crop developmental model: Description and evaluation against measurements of carbon fluxes and carbon stocks at several European agricultural sites, Agr. Ecosyst. Environ., 139, 402–418, 2010.

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, J. Geophys. Res, 106, 7183–7192, 2001.

Thomas, C. K., Law, B. E., Irvine, J., Martin, J. G., Pettijohn, J. C., and Davis, K. J.: Seasonal hydrology explains interannual and seasonal variation in carbon and water exchange in a semi-arid mature ponderosa pine forest in central Oregon, J. Geophys. Res., 114, G04006, doi:10.1029/2009JG001010, 2009.

Thornton, P. E., Running, S. W., and Hunt, E. R.: Biome-BGC: Terrestrial Ecosystem Processes Model, Version 4.1.1, Oak Ridge National Laboratory Distributed Active Archive Center, doi:10.3334/ORNLDAAC/805, 2005.

Tian, H. Q., Chen, G., Liu, M., Zhang, C., Sun, G., Lu, C., Xu, X., Ren, W., Pan, P., and Chappelka, A.: Model estimates of ecosystem net primary productivity, evapotranspiration and water use efficiency in the Southern United States during 1895–2007, Forest Ecol. Manag., 259, 1311–1327, 2010.

Torrence, C. and Compo, G. P.: A practical guide to wavelet analysis, B. Am. Meteorol. Soc., 79, 61–78, 1998.

Torrence, C. and Webster, P.: Interdecadal changes in the ENSO-Monsoon system, J. Climate, 12, 2679–2690, 1999.

Urbanski, S. P., Barford, C., Wofsy, S., Kucharik, C. J., Pyle, E. H., Budney, J., McKain, K., Fitzjarrald, D., Czikowsky, M. J., and Munger, J. W.: Factors controlling CO_2 exchange on timescales from hourly to decadal at Harvard Forest, J. Geophys. Res., 112, G02020, doi:10.1029/2006JG000293, 2007.

Van Gorsel, E., Berni, J. A. J., Briggs, P., Cabello-Leblic, A., Chasmer, L., Cleugh, H. A., Hacker, J., Hantson, S., Haverd, V., Hughes, D., Hopkinson, C., Keith, H., Kljun, N., Leuning, R., Yebra, M., and Zegelin, S.: Primary and secondary effects of climate variability on net ecosystem carbon exchange in an evergreen Eucalyptus forest, Agr. Forest Meteorol., doi:10.1016/j.agrformet.2013.04.027, 2013.

Vargas, R., Detto, M., Baldocchi, D. D., and Allen, M. F.: Multiscale analysis of temporal variability of soil CO_2 production as influenced by weather and vegetation, Glob. Change Biol., 16, 1589–1605, doi:10.1111/j.1365-2486.2009.02111.x, 2010.

Vargas, R., Sonnentag, O., Abramowitz, G., Carrara, A., Chen, J. M., Ciais, P., Correira, A., Keenan, T., Kobayashi, H., Ourcival, J–M., Papale, D., Pearson, D., Pereira, J. S., Piao, S. L., Rambal, S., and Baldocchi, D. D.: Drought influences the accuracy of simulated ecosystem fluxes: a model–data meta–analysis for Mediterranean oak woodlands, Ecosystems, 16, 749–764, doi:10.1007/s10021-013-9648-1, 2013.

Venema, V., Ament, F., and Simmer, C.: A stochastic iterative amplitude adjusted Fourier transform algorithm with improved accuracy, Nonlinear Proc. Geoph., 13, 321–328, 2006.

Verma, S. B., Dobermann, A., and Cassman, K. G.: Annual carbon dioxide exchange in irrigated and rainfed maize-based agroecosystems, Agr. Forest Meteorol., 131, 77–96, 2005.

Wang, W., Dungan, J., Hashimoto, H., Michaelis, A. R., Milesi, C., Ichii, K., and Nemani, R. R.: Diagnosing and assessing uncertainties of terrestrial ecosystem models in a multi-model ensemble experiment: 2. Carbon balance, Glob. Change Biol., 17, 1367–1378, doi:10.1111/j.1365-2486.2010.02315.x, 2010.

Wang, Y. P., Kowalczyk, E. A., Leuning, R., Abramowitz, G., Raupach, M., Pak, B. C., van Gorsel, E., and Luhar, A.: Diagnosing errors in a land surface model (CABLE) in the time and frequency domains, J. Geophys. Res., 116, G01034, doi:10.1029/2010JG001385, 2011.

Weng, E. and Luo, Y.: Soil hydrological properties regulate grassland ecosystem responses to multifactor global change: A modeling analysis, J. Geophys. Res., 113, G03003, doi:10.1029/2007JG000539, 2008.

Williams, M., Schwarz, P. A., Law, B., Irvine, J., and Kurpius, M. R.: An improved analysis of forest carbon dynamics using data assimilation, Glob. Change Biol., 11, 89–105, 2005.

Williams, M., Richardson, A. D., Reichstein, M., Stoy, P. C., Peylin, P., Verbeeck, H., Carvalhais, N., Jung, M., Hollinger, D. Y., Kattge, J., Leuning, R., Luo, Y., Tomelleri, E., Trudinger, C. M., and Wang, Y. -P.: Improving land surface models with FLUXNET data, Biogeosciences, 6, 1341–1359, doi:10.5194/bg-6-1341-2009, 2009.

Williamson, T. B., Price, D. T., Beverley, J. L., Bothwell, P. M., Frenkel, B., Park, J., and Patriquin, M. N.: Assessing potential biophysical and socioeconomic impacts of climate change on forest-based communities: a methodological case study, Natural Resources Canada, Canadian Forest Service, Edmonton, ABInf. Rep. NOR-X-415E, 2008.

Zhan, X. W., Xue, Y. K., and Collatz, G. J.: An analytical approach for estimating CO_2 and heat fluxes over the Amazonian region, Ecol. Model., 162, 97–117, 2003.

Zhou, X. L., Peng, C. H., Dang, Q. L., Sun, J. F., Wu, H. B., and Hua, D.: Simulating carbon exchange in Canadian boreal forests I: model structure, validation, and sensitivity analysis, Ecol. Model., 219, 287–299, 2008.