Environmental Management
DOI 10.1007/s00267-014-0437-1

Relative Linkages of Canopy-Level CO_2 Fluxes with the Climatic and Environmental Variables for US Deciduous Forests

Khandker S. Ishtiaq · Omar I. Abdul-Aziz

Received: 9 April 2014 / Accepted: 18 December 2014

Abstract We used a simple, systematic data-analytics approach to determine the relative linkages of different climate and environmental variables with the canopy-level, half-hourly CO_2 fluxes of US deciduous forests. Multivariate pattern recognition techniques of principal component and factor analyses were utilized to classify and group climatic, environmental, and ecological variables based on their similarity as drivers, examining their interrelation patterns at different sites. Explanatory partial least squares regression models were developed to estimate the relative linkages of CO_2 fluxes with the climatic and environmental variables. Three biophysical process components adequately described the system-data variances. The 'radiation-energy' component had the strongest linkage with CO_2 fluxes, whereas the 'aerodynamic' and 'temperature-hydrology' components were low to moderately linked with the carbon fluxes. On average, the 'radiation-energy' component showed 5 and 8 times stronger carbon flux linkages than that of the 'temperature-hydrology' and 'aerodynamic' components, respectively. The similarity of observed patterns among different study sites (representing gradients in climate, canopy heights and soil-formations) indicates that the findings are potentially transferable to other deciduous forests. The similarities also highlight the scope of developing parsimonious data-driven models to predict the potential sequestration of ecosystem carbon under a changing climate and environment. The presented data-analytics provides an objective, empirical foundation to obtain crucial mechanistic insights; complementing process-based model building with a warranted complexity. Model efficiency and accuracy ($R^2 = 0.55$–0.81; ratio of root-mean-square error to the observed standard deviations, RSR = 0.44–0.67) reiterate the usefulness of multivariate analytics models for gap-filling of instantaneous flux data.

Keywords Relative linkages · CO_2 fluxes · Data-analytics · Pattern recognition · Empirical modeling · Deciduous forests

Introduction

Terrestrial ecosystems are the major components of earth's carbon cycle and traditionally regarded as the reservoirs of green carbon. Turbulent land-atmospheric fluxes of vertical CO_2 contribute to ecosystem-scale carbon budget by balancing between the above-ground assimilatory processes (photosynthesis) and both above-ground and below-ground respiratory processes (Heimann and Reichstein 2008; Cao and Woodward 1998; Schimel et al. 2001). The carbon flux-climate feedback process is sensitive to various meteorological, hydrological, and ecological variables at different spatiotemporal scales. Although much research has been conducted to characterize the role of key environmental stressors and climatic variability on land-atmospheric carbon and heat fluxes (e.g., Schmidt et al. 2011), a robust understanding and prediction of turbulent carbon flux dynamics is

K. S. Ishtiaq
Ecological and Water Resources Engineering Laboratory (EWREL), Department of Civil and Environmental Engineering, Florida International University, 10555 W Flagler Street, EC-3781, Miami, FL 33174, USA
e-mail: kisht001@fiu.edu

O. I. Abdul-Aziz (✉)
Department of Civil and Environmental Engineering, Florida International University, 10555 W Flagler Street, EC-3606, Miami, FL 33174, USA
e-mail: oabdulaz@fiu.edu

Published online: 08 January 2015

yet to be achieved (Piao et al. 2013; Morales et al. 2005; Geider et al. 2001). Investigation of the relative effects of climatic and environmental variations on turbulent carbon fluxes is, therefore, an area of active research. In particular, quantification of the relative linkages of climatic and environmental variables with the vertical CO_2 fluxes is an important step toward building relatively simple empirical to complex mechanistic (i.e., process-based) models for robust predictions and management of ecosystem carbon. For example, altering land uses/cover can change surface albedo (light reflectance), net radiation, temperature, soil moisture, heat fluxes, canopy height, roughness height, and friction velocity. Knowledge and insights on the relative carbon flux linkages of climatic and environmental variables would, therefore, help in developing land management strategies and priorities for a maximum uptake (sequestration) and minimum emission of ecosystem carbon.

Numerous process-oriented biosphere models are available for simulation and prediction of carbon fluxes by mechanistically capturing the necessary ecosystem constituents such as plant photosynthesis, respiration, and soil biogeochemical processes. Examples of mechanistic models include IFUSE (Desai 2010), Ecosystem Demography (ED) (Moorcroft et al. 2001), BIOME-BGC (Running and Coughlan 1988; Running and Gower 1991), Equilibrium Boundary Layers (EBL) (Shir and Bornstein 1977), Carbon Tracker (CT) (http://carbontracker.noaa.gov), HYBRID (White et al. 1999), ECOSYS (Grant et al. 2012), CENTURY (Glimanov et al. 1997), LINKAGE (Post and Pastor 1996), etc. Most models use photosynthetically active radiation, vegetation index, atmospheric CO_2 concentration, air and soil temperatures, vapor pressure deficit, and soil moisture as inputs (Chen et al. 2011; Sims et al. 2008; Schubert et al. 2012; Li et al. 2007; Turner et al. 2006). These models attempt to predict sub-daily to seasonal and inter-annual variability of carbon flux dynamics.

Process-based models are built on numerous scientific hypotheses and their outcome inherently depends on the embedded process formulations and parameterizations (Beer et al. 2010; Keenan et al. 2012). Further, mechanistic modeling generally involves a complex structure; requires data for many input variables that are not always available, and involves a large parameter set making the model predictions quite uncertain. Application of the mechanistic models also requires high computational resources, expert knowledge and specialized skills, hindering their wide-spread applications as tools for a sustainable management of ecosystem carbon under a changing climate and environment.

Data-driven analytic approaches can lay out the foundation for building an appropriate process-based model of warranted complexity. Standard techniques (e.g., linear regressions, nonlinear regressions using mainly Arrhenius and Michaelis–Menten equations, artificial neural networks) have already been successfully applied to fill gaps in measured data of ecosystem fluxes and environmental variables (Falge et al. 2001; Carrara et al. 2003; Lloyd and Taylor 1994; Richardson et al. 2006; Hollinger et al. 2004; Barr et al. 2004; Gove and Hollinger 2006; Aubinet et al. 1999; Braswell et al. 2005; Hui et al. 2004; Stauch and Jarvis 2006). Multi-scale models were also developed to estimate carbon fluxes from selected ecosystems using remotely sensed and in situ observations of geographical, environmental, and meteorological variables (e.g., Jahan and Gan 2013; Wylie et al. 2007; Mäkelä et al. 2008; Byrne et al. 2005; Oechel et al. 2000). However, lack of knowledge on the general pattern of dominant predictors and their complex interactions, as well as site-specificity of estimated model parameters, still pose as major gaps in data-driven, robust predictions of terrestrial carbon fluxes.

In this paper, we present a simple, systematic data-analytics approach to analyze observational data and determine the relative linkages of different climate and environmental variables with the canopy-level, vertical CO_2 fluxes of eight US deciduous forests. Multivariate pattern recognition techniques such as the principal component analysis (PCA) and factor analysis (FA), in concert with the Pearson correlation analysis, were utilized to classify and group climatic, environmental, and ecological variables based on their similarity as drivers, examining their interrelation patterns and relative influences in different forest sites. Explanatory partial least squares regression (PLSR) models were then developed to estimate the relative linkages of CO_2 fluxes with the climatic and environmental variables. Our findings will guide the development of parsimonious empirical models, while informing the building of appropriate mechanistic models, for robust predictions of ecosystem carbon fluxes from the deciduous forests and other similar ecosystems.

Materials and Methods

Study Sites

The study sites include the following eight deciduous broadleaf forests, ranging from the upper mid-western to the northeastern USA (Fig. 1): (i) Bartlett Experimental Forest, New Hampshire; (ii) Harvard Forest, Massachusetts; (iii) Missouri Ozark, Missouri; (iv) Morgan Monroe State Forest, Indiana; (v) Ohio Oak Openings, Ohio; (vi) Silas Little Experimental Forest, New Jersey; (vii) University of Michigan Biological Station (UMBS), Michigan; and (viii) Willow Creek Forest, Wisconsin. Although these sites have similarity in vegetation cover, they represent diversity in climate, topography, land uses, soil and hydrologic patterns, etc. Consideration of the relatively large geographical region, therefore, incorporates a potentially high spatial gradient in land-atmospheric carbon flux dynamics.

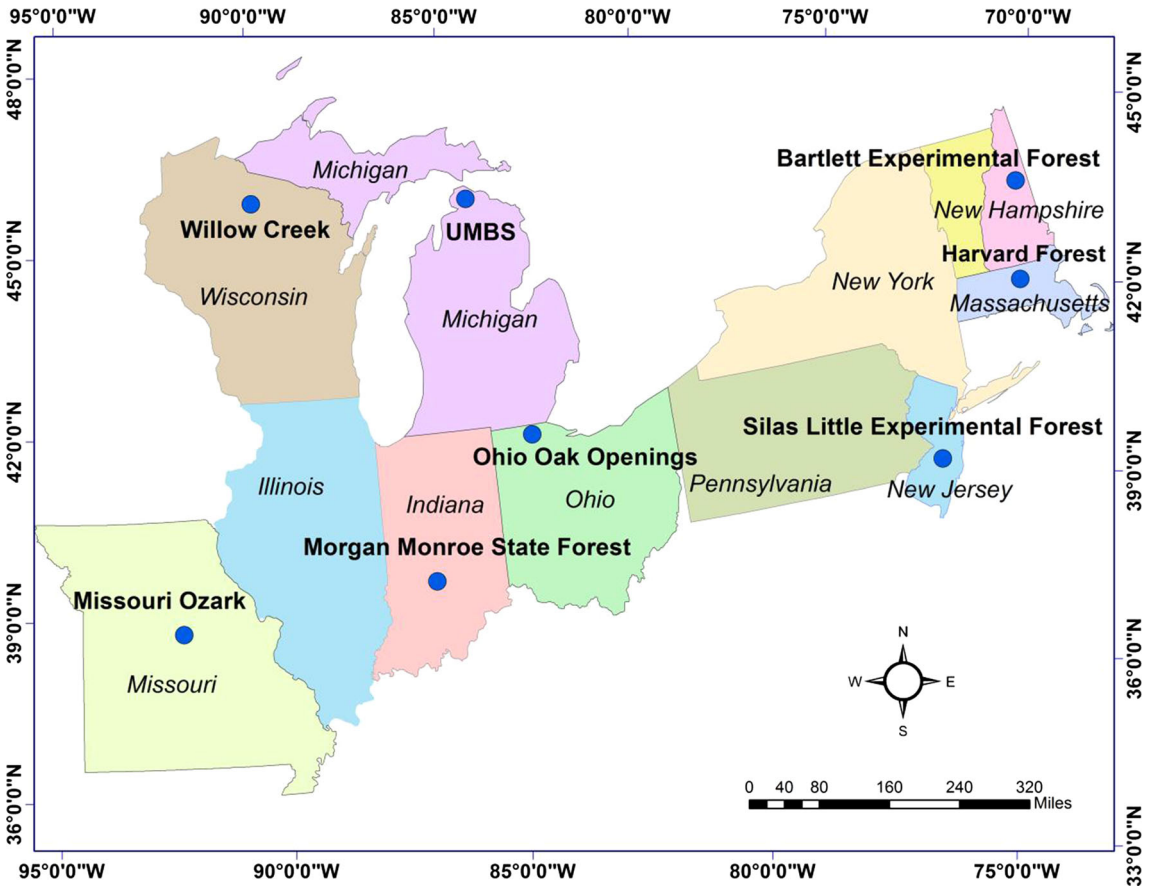

Fig. 1 Locations of the selected eight deciduous forest study sites

The Bartlett Experimental Forest (canopy height: 19 m) has a rolling to mountainous topography with gentle slope and spodosol soil order; climate is nearly extreme with a minimum daytime temperature of around 90 °F during summer and −30 °F during winter. Harvard forest (canopy height: 23 m) has a temperate climate and a sandy soil formation in the top layer and sand-gravel-silt at the bottom. Missouri Ozark (canopy height: 24.2 m) and Morgan Monroe State Forests (canopy height: 27 m) share a similar temperate continental climate, while representing different soil formations (silty loam with a rocky, thin soil cover at Missouri, whereas a combination of clay, loam, and limestone residue at the Morgan site). Ohio Oak Forest (canopy height: 24 m) has a humid climate and a flat terrain with sandy mixed soil. Silas Little Experimental Forest (canopy height: 9.52 m) has a cool, temperate climate, and a sandy soil formation. The UMBS site (canopy height: 12 m) is characterized by the temperate northern climate and a sandy, Entic Haplorthod, glacial till soil. The Willow Creek site (canopy height: 24.2 m) represents the northern continental climate with short, moist growing seasons and cold winters, and a soil formation similar to that of the UMBS site. Overall, the selected eight study sites incorporate the potential effects of a variable canopy height (9.52–27 m), different climatic regimes (humid to temperate to nearly extreme), and diverse soil morphology on the carbon flux dynamics.

Data Sets

Observational data of the half-hourly, canopy-level CO_2 fluxes, and the corresponding climate and environmental variables were obtained for different, recent annual cycles (2006–2011) for the study sites from the AmeriFLUX network (http://ameriflux.ornl.gov) (Table 1); selection of the 5-year period incorporated the effect of multi-year temporal gradients on carbon flux dynamics. Selected variables were vertical CO_2 flux (measured above the canopy without correcting for underlying storage and advection) (F_{CO_2}; denoted as F_C in AmeriFLUX); ambient CO_2 concentration (C_{CO_2}); net radiation (RN); incoming photosynthetically active radiation (PAR); sensible heat flux (SHF); latent heat flux (LHF) (ecosystem water exchanges); air temperature (TA); soil temperature (TS); vapor pressure deficit (VPD); soil water content (SWC); wind speed (WS); and friction velocity (UST). The TA,

Table 1 Means and standard deviations (in parenthesis) of the observed carbon fluxes, and climate and environmental variables for the eight broad-leaf deciduous forest sites of USA

Site	F_{CO_2} ($\mu mol/m^2/s$)	UST (m/s)	TA (°C)	WS (m/s)	SHF (W/m^2)	LHF (W/m^2)	TS (°C)	C_{CO_2} ($\mu mol/mol$)	VPD (kPa)	SWC (%)	RN (W/m^2)	PAR ($\mu mol/m^2/s$)
Bartlett Forest, NH	−3.34 (8.18)	0.44 (0.29)	11.7 (8.90)	1.7 (0.96)	49.00 (112.4)	41.58 (72.38)	9.86 (5.3)	383.4 (17.5)	0.51 (0.51)	35.69 (8.9)	143.40 (224.5)	457.00 (544.7)
Year: 2009; $N = 5{,}597$												
Harvard Forest, MA	−1.37 (7.89)	0.53 (0.30)	8.71 (8.69)	2.19 (0.95)	36.04 (120.48)	30.17 (63.03)	8.61 (6.25)		0.41 (0.39)		95.85 (198.86)	316.10 (481.0)
Year: 2006; $N = 4{,}446$												
Missouri Ozark, MO	−1.43 (6.71)	0.435 (0.30)	12.97 (9.97)	2.74 (1.29)	33.28 (92.01)	54.61 (92.73)	12.69 (7.31)	390.90 (15.27)	0.635 (0.55)	43.12 (7.91)	107.40 (233.69)	411.10 (599.4)
Year: 2009; $N = 13{,}194$												
Morgan Forest, IN	−1.51 (6.63)	0.46 (0.27)	12.29 (11.9)	3.54 (1.28)	32.46 (88.71)	42.86 (90.80)			0.71 (0.65)	35.14 (11.48)	99.45 (221.57)	360.50 (526.7)
Year: 2010; $N = 7{,}080$												
Ohio Oak, OH	−4.47 (10.49)	0.49 (0.32)	11.10 (10.6)	2.41 (1.14)	48.48 (106.34)	66.45 (108.8)	11.68 (5.66)	384.1 (16.55)	0.61 (0.59)	6.13 (1.58)	132.60 (236.83)	500.50 (634.5)
Year: 2009; $N = 7{,}729$												
Silas Forest, NJ	−0.44 (6.07)	0.41 (0.28)	12.69 (10.4)	1.82 (1.08)	36.44 (93.90)	56.29 (109.9)	12.91 (7.41)		0.67 (0.69)		84.43 (194.25)	336.14 (495.95)
Year: 2011; $N = 13{,}188$												
UMBS Forest, MI	−1.76 (6.97)	0.49 (0.32)	9.85 (11.0)	3.67 (1.65)	38.67 (94.98)	44.10 (89.95)	10.14 (7.10)	389.5 (11.76)	0.49 (0.48)	7.81 (2.41)	113.20 (209.67)	406.6 (539.47)
Year: 2011; $N = 11{,}688$												
Willow Creek, WI	−2.21 (6.99)	0.55 (0.29)	7.38 (11.1)	2.52 (1.09)	41.38 (93.70)	40.84 (81.62)	8.87 (7.03)	389.4 (13.00)	0.24 (0.41)	30.86 (4.32)	110.50 (214.21)	398.6 (550.2)
Year: 2006; $N = 8{,}879$												

"Year" and "*N*" refer to data collection year and sample size, respectively. "Blank" cells in the table indicate the "missing data". F_{CO_2}, UST, TA, WS, SHF, LHF, TS. C_{CO_2}, VPD, SWC, RN, and PAR refer, respectively, to CO_2 flux, friction velocity, air temperature, wind speed, sensible heat flux, latent heat flux, soil temperature, ambient CO_2 concentration, vapor pressure deficit, soil water content, net radiation, and photosynthetically active radiation

WS, PAR, LHF, and SHF represent measurements just above the canopy level. The radiation measurements reflect the impacts of cloudiness, which can significantly control plant photosynthesis (Fuentes and Wang 1999).

Choice of the selected study years (2006–2011) was based on the most recent, relatively high availability of reliable data; the set of participatory variables was determined by a preliminary data analysis and leveraging current understanding of carbon flux dynamics in the selected ecosystems. The data sets contain four distinct process partitions of flux, radiation, above-ground and below-ground environmental variables. Most of the process-based carbon flux prediction models used these components for system representation (Carvalhais et al. 2005; Desai 2010; Chen et al. 2011); some empirical models also adopted a similar approach (e.g., Wylie et al. 2007).

Collected data are classified as level-2, which passed through QA/QC checks. Since data sample for each forest site was very large, gaps in the data matrix were not filled in order to avoid any additional biases in the empirical model building. We applied a two-step procedure to remove the unsuitable data and prepare the final data set for each of our eight sites. First, half-hourly data panels representing gaps for more than two participatory variables were removed from further analysis. This procedure led to the exclusion of half-hourly panel data over a year by around 24 % for both Missouri Ozark and Silas Forests, by 33 % from the UMBS, by 49 % from the Willow Creek, by 55 % from the Ohio Oak Forest, by 59 % from the Morgan Forest, by 67 % from the Bartlett Forest, and by 74 % from the Harvard Forest (see Table 1 for the final sample sizes). Although according to Falge et al. (2001), flux towers can encounter around 35 % missing data annually at the half-hourly scale, our removal percentages were higher for most stations because some variables (e.g., soil water content), which had not been collected by the eddy covariance method, encountered higher gaps than the flux data. Second, the gap-filtered data for each variable at each station were plotted with time (not shown) to visually check the presence of any unreasonable spikes (i.e., outliers), which were subsequently removed. This secondary filtering led to the removal of 1 half-hourly observation panel for the Silas Forest, 3 observation panels for the UMBS, 2 observation panels for the Willow Creek, 3 observation panels for the Harvard Forest, and 14 observation panels for the Bartlett Forest. The final data sets ($N = 4{,}446$–$13{,}194$; see Table 1) incorporated the effects of analyzing an equivalent single to multiple seasons on carbon flux dynamics among different stations. Per AmeriFLUX sign conventions, positive sign represents upward fluxes (land/forest to atmosphere) of F_{CO_2}, LHF and SHF; negative refers to their downward fluxes (atmosphere to land/forest). In contrast, positive values of radiation (RN and PAR) indicate downward (atmosphere to forest) fluxes and vice versa.

Although photosynthesis and respiration are two different processes and can be dictated by different drivers (e.g., radiation vs. temperature), they often share common stressors. Partitioning of F_{CO_2} into photosynthesis and respiration was not considered in this study because the objective was to understand (and quantify) the relative linkages and groupings of different variables influencing the overall diurnal (24-h) cycle of carbon exchanges. Since the study goal was to determine the relative linkages of mainly climate and environmental variables with the canopy-level vertical carbon fluxes, biological variables such as the canopy leaf-area-index (LAI) were not included in the data matrices. Precipitation was not also included because of the lack of availability of 30-min interval; instead SWC was used as a surrogate variable to represent hydrologic effect. We also considered latent and sensible heat fluxes in the data matrix in order to directly incorporate ecohydrologic dynamics.

Data Analysis and Empirical Modeling

We used a simple data-analytics approach that incorporates Pearson correlation analysis, as well as the multivariate PCA, FA, and PLSR to emphasize the entire methodology and overall outcomes, rather than the individual analyses or analysis steps. The data-analytics methodology is briefly summarized in a flow diagram (Fig. 2); the analyses and modeling were done using multiple computation software and programming platforms such as MATLAB, R, and Microsoft Excel.

Pearson Correlation Analysis

Pearson product-moment correlation coefficients were computed to obtain background information on the linear dependency between F_{CO_2} and the climatic/environmental variables for the different deciduous forest sites. Triangular correlation matrices involving all participatory variables were also computed to obtain a preliminary understanding on the multicollinear structure of the data matrix.

Principle Component Analysis (PCA) and Factor Analysis (FA)

Multivariate advanced data reduction, interpretation and pattern recognition techniques such as the PCA and FA mine relationships among participating variables in a matrix of independent factors. PCA explains the variance–covariance structure of a data matrix through some orthogonal, linear combinations of original variables, emphasizing data

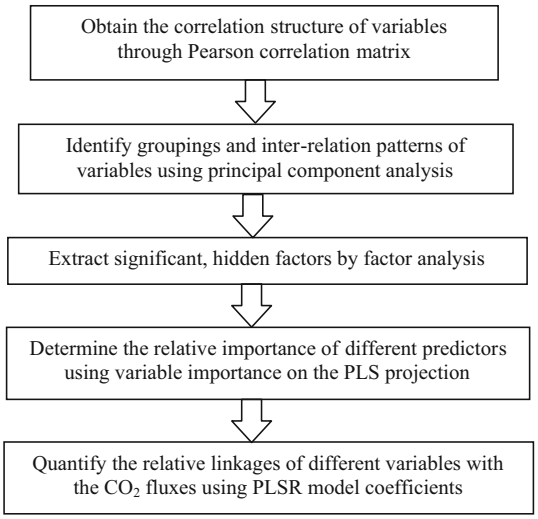

Fig. 2 The data-analytics methodology to determine the relative carbon flux linkages of different climate and environmental variables

interpretation, and reduction in an unsupervised manner (Peres-Neto et al. 2003; Jolliffe 1993; Mahbub et al. 2010). It was previously applied to investigate the spatial representativeness of AmeriFLUX network by grouping homogeneous areas (Hargrove and Hoffman 2005). A virtue of PCA is that it can unravel relationships hidden in the original data and allows interpretations that are not easy to make using a Pearson correlation matrix. We applied explanatory PCA on the data matrices of the biological, climatic, and environmental variables for the eight study sites. In order to bring different variable units and data sources on a comparable reference scale, data for all the variables were standardized (and made dimensionless to obtain *Z*-scores) by dividing their instantaneous deviations from the corresponding annual averages by the respective standard deviations (i.e.,$Z = \frac{X-\mu}{\sigma}$; where $Z = Z\text{-score} =$ normalized variable, $X =$ original variable, $\mu =$ annual average of X, and $\sigma =$ standard deviation of X). First two principal components (PCs) were extracted from the loading matrix (that represents correlation between the PCs and the original variables) and displayed through biplots, which exhibit the possible groupings and interrelations (orientation and correlation strengths) among participatory variables.

FA characterizes the covariance liaisons among many variables with a few rudimental, but unobservable, quantities called factors. FA has been successfully applied for data mining and analysis in many disciplines (Panda et al. 2006; Dragon 2006; and Liu et al. 2003). We applied FA in order to reanalyze the normalized data and verify findings from the PCA by explaining the system variances with fewer latent variables (factors). Individual latent factors were extracted based on an initial eigenvalue criterion (eigenvalue >1.0). Additionally, the "varimax" orthogonal rotation was performed, maximizing the sum of the variances of the loading matrices to optimize loading (i.e., correlation) values of the different variables on each factor. Factors extracted thereby were able to describe most of the variances of the data matrices for different study sites.

Partial Least Squares Regression (PLSR)

PLSR is a sophisticated data-driven method to integrate features from a supervised principal component analysis and multiple regressions, explaining the linear relationships between the dependent (i.e., response) variable and independent (i.e., predictor) variables (Wold 1966, 1982). The unique advantage of using PLSR over traditional multiple linear regressions is that it largely eliminates high variability and instability of estimated parameters caused by multicollinearity among predictors. Since the PLSR regression is performed in the transformed orthogonal planes using the independent PLS components by maximally linking data covariance with the response variable, all the predictor variables can be kept in the final model (Kuhn and Johnson 2013). The regression coefficients of the optimal PLSR model are leveraged to compute the regression coefficients (BETA) of the original independent variables by inverting the linear transformations between the PLS components and original variables. Since the issue of multicollinearity is resolved in the PLSR domain, the derived regression coefficients (BETA) of the original variables should ideally be unaffected by any multicollinearity existing in the data matrices.

In order to quantify the relative linkages of climatic and environmental variables with the vertical CO_2 fluxes, normalized (dimensionless) PLSR models were developed using Z-scores of all participatory variables. The PLSR models were trained (i.e., fitted) and verified (i.e., tested) with half-hourly data ($N = 4{,}446$–$13{,}194$ among different stations) using the SIMPLS algorithm (de Jong 1993; Hubert and Branden 2003) and a 10-fold cross-validation method (Kuhn and Johnson 2013). The model intercept was "zero" for all sites since *Z*-score variables were used for the model fitting.

Application of PLSR requires the selection of optimum number of PLS components to ensure minimal prediction error (optimum *F* statistics) while retaining model stability. The optimal numbers of PLS components were determined using the Akaike Information Criterion (AIC) (Akaike 1974) and coefficient of determination (R^2), as obtained from the 10-fold cross-validations. Subject to the different sample sizes, a normalized $AIC_{F_{CO_2}}$ was defined to bring all sites to a comparable scale as follows:

$$AIC_{F_{CO_2}}(p) = \ln\left(\frac{SSE_{F_{CO_2}}}{N}\right) + \frac{2p}{N} \quad (1)$$

where p is the total number of the model PLS components, N is the sample size, and $SSE_{F_{CO_2}}$ is the total sum of squared error upon estimation of carbon fluxes (F_{CO_2}). Using the optimum number of PLS components, we employed both the "variable importance in the projection (VIP)" (i.e., PLS-VIP) and "regression coefficients (BETA)" (i.e., PLS-BETA) methods as the complementary approaches to determine the relative importance of different predictors for the model response (Wold et al. 1993, 2001; Chong and Jun 2005). Higher VIP scores and higher regression coefficients indicate more influential predictors in the latent predictor-response matrix; as a rule of thumb, VIP scores exceeding 1.0 can be considered as the most informative predictors for the response (Kuhn and Johnson 2013). Further, the regression coefficients (BETA) of the Z-score PLSR models can represent the relative linkages of the predictor variables (C_{CO_2}, RN, PAR, SHF, LHF, TA, VPD, TS, WS, UST, and SWC) with the response variable (F_{CO_2}).

Results

Correlation Structure of the Data

The Pearson correlation coefficients between half-hourly F_{CO_2} and the corresponding climatic/environmental variables were significant at the 95 % level of confidence ($\alpha = 0.05$ for a two tailed test) (Table 2). Since the participatory variables (e.g., carbon and heat fluxes, temperatures) had both positive and negative values, absolute values of the correlation coefficients were used to describe the linear correspondences of different variables. For all the study sites, F_{CO_2} demonstrated strong linear correspondences with the biosphere radiations (RN and PAR) ($|r| = 0.61$–0.71) and LHF (ecosystem-atmospheric heat and water exchanges due to evapotranspiration) ($|r| = 0.65$–0.87); while showing moderate correlations ($|r| = 0.29$–0.56) with the SHF (ecosystem-atmospheric heat exchanges due to temperature gradient), as well as with the ambient carbon storage (C_{CO_2}) ($|r| = 0.46$–0.57). Moderate linear correspondences of F_{CO_2} were also apparent with the air temperature (TA) ($|r| = 0.28$–0.51) and soil temperature (TS) ($|r| = 0.24$–0.51) among the different study sites. Vapor pressure deficit (VPD) and available soil moisture (SWC) showed weak to moderate correlations with F_{CO_2} ($|r| = 0.10$–0.54 and $|r| = 0.01$–0.33, respectively). In contrast, the linear correspondences of F_{CO_2} with the aerodynamic drivers (WS and UST) were relatively weak ($|r| = 0.03$–0.24). Furthermore, the triangular correlation matrices (not shown) for different study sites revealed high mutual correlations among the flux related variables (SHF, LHF, RN, and PAR). For example, the correlation coefficient (r) between RN and PAR was 0.99; correlation of the radiation variables (PAR and RN) with the SHF and LHF ranged, respectively, from 0.80 to 0.87 and from 0.71 to 0.80. The temperatures variables (TS, TA, and VPD), as well as the velocity variables (WS and UST), were also notably correlated within each group. This indicates the presence of a substantial multicollinearity in the data matrix of the climatic and environmental variables.

Dominant Groups and Orientation of the Variables

For the eight study forests, the first two PCs explained from 61.95 to 75.17 % of the total data variances exhibited by the participatory (climatic, environmental, and biological) variables. PCA loading matrices (showing the correlation coefficients between the PCs and the original variables) for different stations are presented through biplots (Figs. 3, 4).

The first two PCs explained 61.95 % of total data variances for the Bartlett Experimental Forest (Fig. 3a). The orientations and lengths of SHF, PAR, RN, VPD, and LHF suggest strong interrelationships, forming a dominant group (A) that highly correlates with F_{CO_2} and C_{CO_2}. TA and TS formed a second group (B), which appears to be moderately correlated with F_{CO_2}. SWC, WS and UST formed the third group (C) that is relatively orthogonal to (i.e., weakly correlated with) F_{CO_2}. Nearly orthogonal orientations of groups A, B, and C suggest three different and relatively uncorrelated variance-based clusters of variables hidden in the data;

Table 2 Pearson correlation coefficients between the carbon fluxes (F_{CO_2}) and the climate/environmental variables for the eight deciduous forest study sites

Site	RN	PAR	LHF	SHF	C_{CO_2}	TA	TS	VPD	SWC	WS	UST
Bartlett Forest	−0.70	−0.70	−0.72	−0.56	0.46	−0.51	−0.39	−0.10	0.06	0.03	0.03
Harvard Forest	−0.66	−0.67	−0.65	−0.37		−0.39	−0.37	−0.44		−0.03	−0.12
Missouri Ozark	−0.61	−0.61	−0.71	−0.33	0.47	−0.28	−0.24	−0.40	0.17	−0.04	−0.18
Morgan Forest	−0.71	−0.71	−0.84	−0.40		−0.39		−0.38	0.01	0.24	0.06
Ohio Oak	−0.71	−0.70	−0.87	−0.29	0.52	−0.43	−0.33	−0.16	0.26	0.09	0.04
Silas Forest	−0.69	−0.66	−0.77	−0.46		−0.36	−0.26	−0.54		0.05	−0.15
UMBS Forest	−0.67	−0.67	−0.80	−0.44	0.53	−0.38	−0.34	−0.50	0.25	0.07	−0.16
Willow Creek	−0.65	−0.67	−0.84	−0.40	0.57	−0.51	−0.51	−0.42	0.33	−0.20	0.05

Blank indicates missing data

A is dominated by the radiation and flux variables, whereas B and C are dictated by, respectively, the temperature and aerodynamic variables. The group-A variables and C_{CO_2}, due to their strongly non-orthogonal orientations with F_{CO_2}, are likely the dominant predictors of the carbon fluxes; group-B variables would be the moderately strong predictors, while group-C variables could be the weakest predictors. Similar variable-groupings and interrelation patterns were observed for the Harvard Forest (with missing SWC data) (Fig. 3b), UMBS Forest (Fig. 3c), and Missouri Ozark Forest (Fig. 3d), where the first two PCs explained, respectively, 71.26, 68.46, and 63.65 % of their total data variances. Although the quadrant-locations of group-B and C were flipped for the Missouri Forest, compared to the other three stations, the relative orthogonality of the three groups, as well as that between F_{CO_2} and group-C variables, were common for all four sites (Fig. 3).

The Silas Little Forest (with missing SWC data), as well as the Willow Creek Forest, demonstrated a slightly different pattern than that of Bartlett Forest, with the first two PCs explaining, respectively, 75.17 and 67.03 % of the total data variances (Fig. 4a, b). The group-A variables still demonstrated the strongest links with F_{CO_2}; however, they split into two distinct, non-orthogonal groups of A1 (RN, PAR, SHF) and A2 (VPD, LHF). Similar to the Bartlett Forest, the group-B (temperature) variables showed moderate correspondences, whereas the group-C (wind speed and soil water) variables appeared to be weakly interrelated with the F_{CO_2}. The Morgan Monroe State Forest (the first two PCs explained 64.38 % of the total data variance) also showed a slightly different correlation and grouping pattern compared to that of Bartlett or similar forests (Fig. 4c). Although group-C was similar to the previous stations showing a relatively orthogonal orientation with F_{CO_2}, the VPD emerged out of group-A and loaded highly with TA to form a moderately linked group, D; this regrouping could be partly caused by the missing soil temperature data for this station. However, at the Ohio Oak Forest, where the two PCs

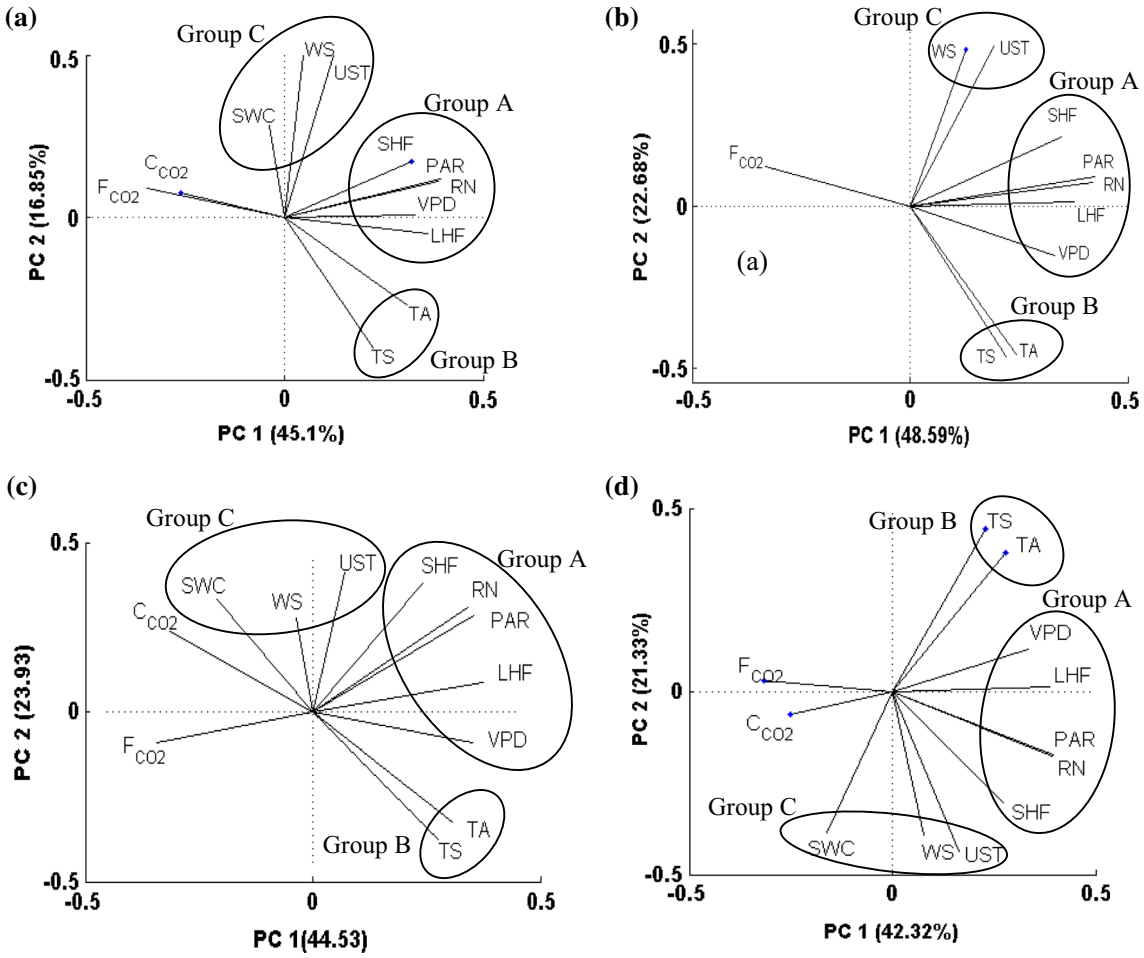

Fig. 3 Biplots obtained from principal component analysis showing the groupings and interrelation patterns of the climate, environmental, and biological variables for (**a**) Bartlett Experimental Forest (NH), (**b**) Harvard Forest (MA), (**c**) UMBS Forest (MI), and (**d**) Missouri Ozark Forest (MO). Percent variance explained by each PC is shown in *parenthesis*

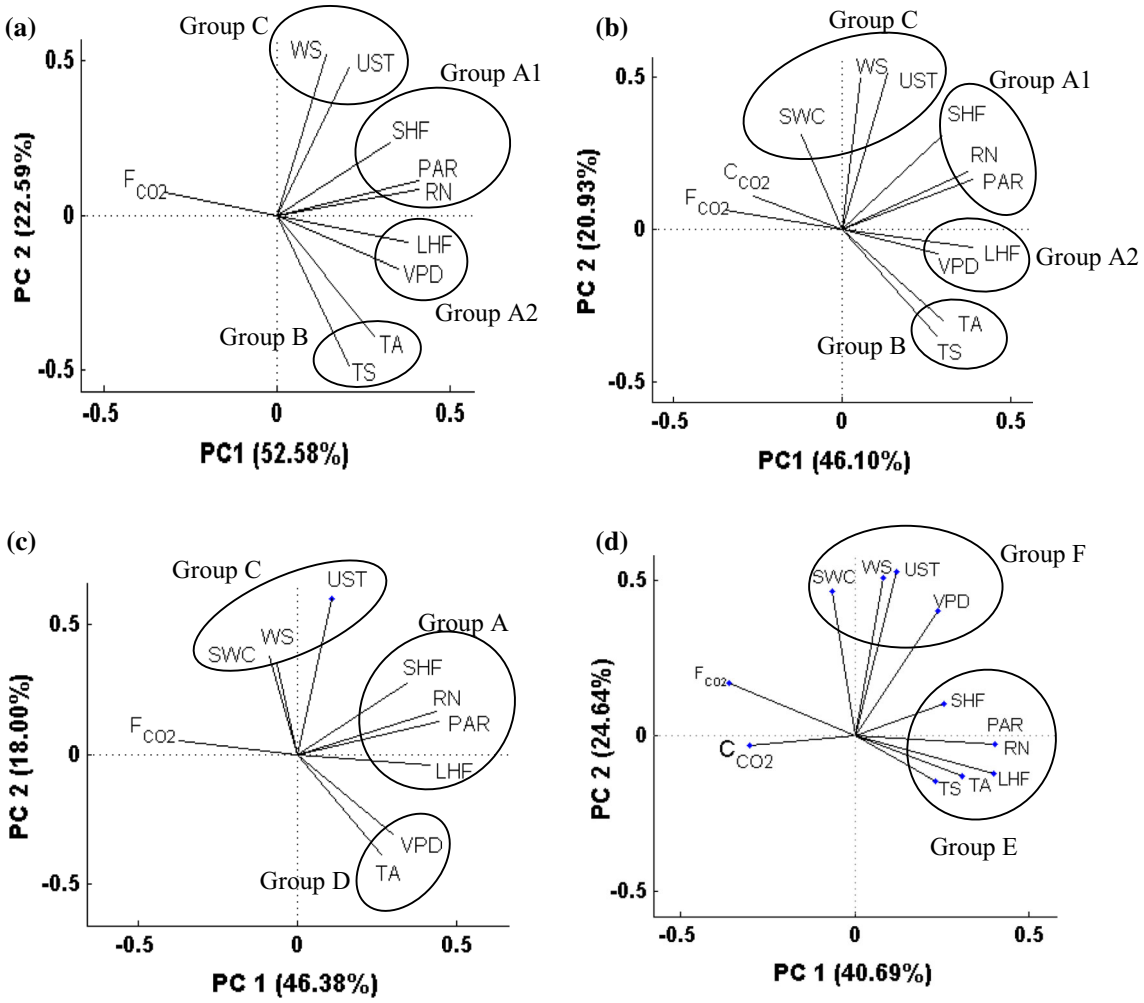

Fig. 4 Biplots obtained from principal component analysis showing the groupings and interrelation patterns of the climate, environmental, and biological variables for (**a**) Silas Little Forest (NJ), (**b**) Willow Creek Forest (WI), (**c**) Morgan Monroe State Forest (IN), and (**d**) Ohio Oak Openings Forest (OH). Percent variance explained by each PC is shown in *parenthesis*

explained 65.33 % of total data variance, all the variables appeared to cluster into two distinct groups of E (SHF, PAR, RN, LHF, TA, and TS) and F (SWC, WS, UST, and VPD) (Fig. 4d). The group-E has a nearly linear (with 180°) orientation with F_{CO_2}, suggesting likely strong linkages between them; the group-F variables were nearly orthogonally oriented with F_{CO_2}, indicating their possibly weak linkages. The ambient carbon storage (C_{CO_2}) was non-orthogonally oriented with F_{CO_2}, suggesting their strong linkage.

Significant Hidden Factors

The eigenvalue criterion (eigenvalue >1) led to three independent latent factors for all eight sites (Table 3). This means that the extracted three factors adequately summarized the system variance at each site; the first factor explained the most variance (40.69–52.6 %), then the second factor (16.8–24.64 %), while the third factor described the least variance (8.9–15.5 %). FA with 'varimax' optimization provided more precise information into the hidden patterns of the data matrices than that shown by the two PCs (see Figs. 3, 4). Since the standardized participatory variables had both positive and negative values, the FA outcomes were interpreted mainly based on the magnitudes (ignoring the positive or negative signs) of the factor loadings.

The radiation and heat fluxes (RN, PAR, SHF, and LHF) generally loaded highly (0.65–0.99) on the first factor, which had moderate to high loadings (−0.57 to −0.70) with F_{CO_2} at different stations. The pressure variable (VPD) showed moderate loadings (0.41–0.53) on Factor-1 for all but one (Morgan State Forest) sites. The near-canopy carbon storage (C_{CO_2}) showed a moderate loading (−0.38) on Factor-1 only for the Bartlett Forest site. The higher loadings of RN, PAR, SHF, and LHF indicate their

Table 3 Dominant latent factors extracted from the data matrices of the eight deciduous forest study sites

Site	Factors	WS	UST	SWC	VPD	TS	TA	C_{CO_2}	SHF	LHF	RN	PAR	F_{CO_2}
Bartlett Forest (NH)	Fac 1	0.02	0.20	−0.01	**0.48**	0.08	0.28	**−0.38**	**0.91**	**0.65**	**0.97**	**0.97**	**−0.65**
	Fac 2	0.04	0.02	−0.14	**0.53**	**0.88**	**0.94**	−0.35	−0.01	0.43	0.20	0.22	−0.38
	Fac 3	**0.99**	**0.89**	0.14	0.22	−0.10	−0.02	−0.04	0.06	0.05	0.09	0.12	0.03
Harvard Forest (MA)	Fac 1	0.08	0.25		**0.51**	0.12	0.16		**0.86**	**0.68**	**0.95**	**0.97**	**−0.63**
	Fac 2	−0.06	−0.06		**0.49**	**0.93**	**0.98**		−0.03	0.27	0.19	0.17	−0.29
	Fac 3	**0.91**	**0.97**		0.08	−0.07	−0.06		0.20	0.19	0.14	0.16	0.02
UMBS Forest (MI)	Fac 1	−0.07	0.27	−0.05	**0.46**	0.03	0.14	−0.22	**0.85**	**0.67**	**0.98**	**0.98**	**−0.63**
	Fac 2	−0.06	−0.10	**−0.60**	**0.64**	**0.96**	**0.96**	**−0.72**	−0.10	**0.42**	0.12	0.16	−0.34
	Fac 3	**0.86**	**0.96**	0.19	0.03	−0.11	−0.07	0.00	0.12	0.07	0.10	0.07	−0.03
Missouri Ozark (MO)	Fac 1	0.03	0.30	−0.07	**0.41**	0.09	0.17	−0.29	**0.83**	**0.71**	**0.97**	**0.97**	**−0.58**
	Fac 2	−0.08	−0.09	**−0.50**	**0.62**	**0.91**	**0.97**	−0.26	−0.09	0.39	0.18	0.19	−0.20
	Fac 3	**0.99**	**0.78**	0.18	0.22	−0.15	0.00	−0.14	0.12	0.08	0.12	0.13	−0.04
Silas Forest (NJ)	Fac 1	0.13	0.29		**0.53**	0.11	0.23		**0.84**	**0.72**	**0.95**	**0.94**	**−0.67**
	Fac 2	−0.09	0.00		**0.60**	**0.91**	**0.97**		0.01	0.37	0.24	0.22	−0.21
	Fac 3	**0.99**	**0.86**		0.15	−0.17	0.01		0.22	0.07	0.19	0.23	0.02
Willow Creek (WI)	Fac 1	0.08	0.28	−0.02	**0.41**	0.18	0.25	−0.34	**0.88**	**0.67**	**0.95**	**0.95**	**−0.57**
	Fac 2	−0.10	−0.07	−0.36	**0.46**	**0.95**	**0.92**	**−**0.39	0.01	**0.51**	0.21	0.26	**−0.45**
	Fac 3	**0.88**	**0.96**	0.16	0.03	−0.12	−0.08	−0.04	0.21	0.07	0.16	0.14	−0.08
Morgan Forest (IN)	Fac 1	−0.21	0.30	0.04	0.34		0.29		**0.82**	**0.78**	**0.99**	**0.98**	**−0.70**
	Fac 2	0.04	−0.14	**−0.44**	**0.88**		**0.75**		0.00	0.32	0.09	0.17	−0.19
	Fac 3	**0.74**	**0.94**	0.01	0.04		−0.09		0.12	−0.05	0.03	0.02	0.13
Ohio Oak (OH)	Fac 1	0.07	0.25	−0.03	**0.52**	0.05	0.25	−0.34	**0.83**	**0.67**	**0.96**	**0.97**	**−0.65**
	Fac 2	−0.15	−0.11	**−0.49**	**0.60**	**0.87**	**0.90**	**−0.44**	−0.14	**0.50**	0.21	0.22	−0.37
	Fac 3	**0.96**	**0.89**	0.22	0.09	−0.24	−0.07	−0.13	0.25	0.04	0.13	0.11	0.04

"Fac" represents factor. Factor 1: Radiation-energy factor; Factor 2: Temperature-hydrology factor; Factor 3: Aerodynamic factor. Bold values indicated variables having moderate to high loadings (correlations) on factors. "Blank" refers to the missing data

dictating role on the first hidden factor, which, therefore, can be termed 'radiation-energy' factor. Temperature variables (TA and TS) loaded highly (0.75–0.98) with Factor-2, which showed low to moderate loadings (−0.19 to −0.45) with F_{CO_2}. The VPD had moderate to high loadings (0.46–0.88) on Factor-2 for all stations; the soil hydrology variable (SWC) loaded moderately (−0.44 to −0.50) on Factor-2 for the Missouri Ozark, Morgan State, and Ohio Oak Forests, while loading slightly highly (−0.60) for the UMBS Forest. The C_{CO_2} showed moderate to high loadings (−0.44 to −0.72) on Factor-2 for the Ohio Oak and UMBS sites only; of the radiation and heat fluxes, only LHF loaded moderately (0.42–0.51) on Factor-2 for the UMBS, Ohio Oak, and Willow Creek Forests. Since temperature related variables and SWC dominated Factor-2, it was termed 'temperature-hydrology' factor. Factor-3 (termed 'aerodynamic' factor) had a poor loading (0.02–0.13) with F_{CO_2}; only the WS and UST showed notable loadings (0.74–0.99) on Factor-3 at different sites.

Similar to the PCA outcomes, the highest loadings of F_{CO_2} with Factor-1 (which explained most of the system variances) refer to the relatively strong linkages of radiation and heat fluxes with the turbulent, vertical CO_2 fluxes; the strong loadings of LHF with both Factor-1 and Factor-2 (ranking second in explaining system variances and F_{CO_2} loadings) indicate a dictating linkage of LHF with F_{CO_2}. Notable double-factor (Factor-1 and 2) associations of VPD and C_{CO_2} also suggest their relatively strong linkages with the vertical carbon fluxes. The very high loadings of TA and TS on Factor-2, as well as that of WS and UST on Factor-3, can indicate their appreciable linkages with F_{CO_2}. The moderate to strong loadings of multiple variables on the 'radiation-energy' and 'temperature-hydrology' factors reiterate the interrelations (i.e., collinearity) among the radiation, heat fluxes, and temperature variables.

Relative Carbon Flux Linkages of the Climate and Environmental Variables

A combination of AIC and R^2 criteria, obtained through a 10-fold cross-validation method, showed that a total of 3–5 PLS components led to the optimum PLSR models (minimum $AIC_{F_{CO_2}}$ and maximum R^2), whereas three PLS

components captured most variations in F_{CO_2} for different sites (Fig. 5a, b); this is consistent with the FA outcome of three independent factors adequately describing the variation in the overall data matrices (Table 3). Ranges of optimal model fitting efficiency ($R^2 = 0.55$–0.81) and accuracy (ratio of root-mean-square error to the standard deviation of observations, RSR = 0.44–0.67; mean square error, MSE = 0.19–0.45) showed impressive predictions of F_{CO_2} for different study sites (Table 4). The model residuals were approximately normally distributed with constant variances (not shown). The R^2 indicated the amount of observed data variance explained by the model (i.e., model efficiency), whereas the RSR (see notes of Table 4 for the mathematical expression) and MSE indicated the accuracy of model fitting. Moriasi et al. (2007) provided a range of RSR values for the evaluation of model accuracy; an RSR from 0 to 0.50 indicates a perfect to very good model, from 0.5 to 0.6 indicates a good model, and from 0.6 to 0.7 refers to a satisfactory model; a model with RSR >0.70 is considered unsatisfactory.

The variable importance in the PLS projection (VIP) scores and regression coefficients (BETA), as obtained with the optimum number of PLS components, quantified the relative linkage of each predictor with the model response (F_{CO_2}) (Fig. 6; Table 4). Since the *Z*-scores of the participatory variables had both positive and negative values, interpretation of the type (e.g., mutual increase or decrease) of their relative carbon flux linkages based on the positive or negative sign of BETA (β) would be potentially misleading; the predictive influence of individual predictors of F_{CO_2} was, therefore, evaluated by comparing the absolute values of the associated β. Total relative linkages of the 'radiation-energy' (β_{RHF}), 'temperature-hydrology' (β_T), and 'aerodynamic' (β_W) components were computed, respectively, as:

$$\beta_{RHF} = \sqrt{\beta_{RN}^2 + \beta_{PAR}^2 + \beta_{LHF}^2 + \beta_{SHF}^2};$$

$$\beta_T = \sqrt{\beta_{TA}^2 + \beta_{TS}^2 + \beta_{VPD}^2 + \beta_{SWC}^2}; \text{ and}$$

$$\beta_W = \sqrt{\beta_{WS}^2 + \beta_{UST}^2}$$

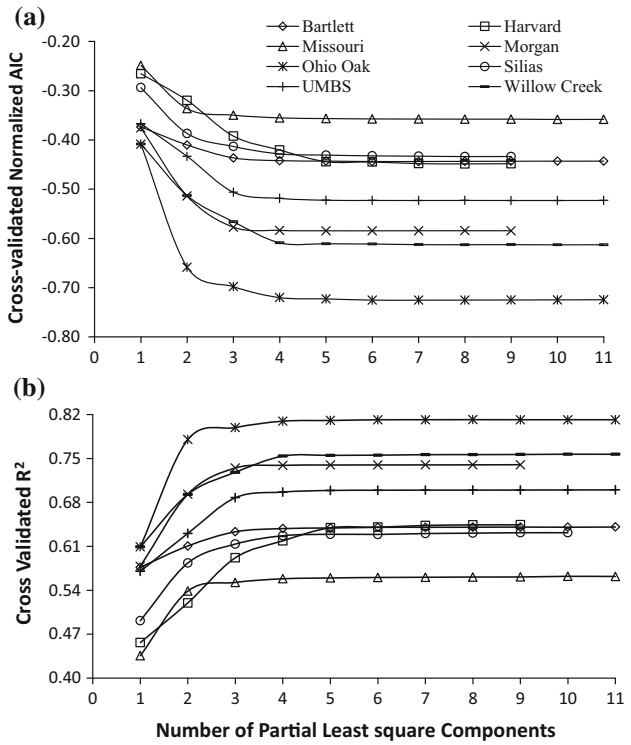

Fig. 5 Plot of cross-validated (**a**) normalized AIC and (**b**) fitting efficiency (R^2) for F_{CO_2} with the number of incorporated partial least squares (PLS) components

Bartlett Experimental Forest

The LHF, RN, PAR, SHF, and VPD had VIP scores higher than unity (1.0) and were potentially strongly linked with F_{CO_2} (Fig. 6a); smaller VIPs of TA, C_{CO_2}, and TS indicated their likely moderate linkages with F_{CO_2}; much lower VIP scores indicated relatively weak linkages of F_{CO_2} with UST, WS, and SWC. Model coefficients (β) revealed almost similar linkage pattern with some exceptions (Table 4); F_{CO_2} showed relatively strong linkage with C_{CO_2} and relatively weak linkages with SHF and TA. LHF was the strongest predictor in regression, exhibiting around 2 times stronger linkages with F_{CO_2} than that of RN, PAR and C_{CO_2}; 2.5–3.5 times stronger linkages than that of VPD and TS; and around 9–36 times stronger linkages than that of the aerodynamic and soil moisture variables (WS, UST, and SWC). Based on the ratio of β_{RHF}/β_T and β_{RHF}/β_W, the 'radiation-energy' component had around 2.5 and 11 times stronger linkages with the carbon fluxes than that of, respectively, the 'temperature-hydrology' and 'aerodynamic' components.

Harvard Forest

The VIP scores and PLSR coefficients slightly differed for this site. The radiation and heat flux variables of PAR, RN, LHF, and SHF had the higher VIP scores, referring to their strong linkages with F_{CO_2} (Fig. 6b). Lower VIP scores of the temperature and aerodynamic variables (VPD, TA, TS, WS, and UST) indicated their relatively weak to moderate linkages with F_{CO_2}. In terms of regression coefficients (β; Table 4), F_{CO_2} had relatively high linkages with SHF, RN, PAR, and LHF; moderate linkages with VPD and UST; and low linkages with TA, TS, and WS. Unlike other study sites, SHF appeared to be the strongest predictor of F_{CO_2}, although VIP scores indicated stronger dominance of LHF over SHF. Overall, the 'radiation-energy' group had approximately 7

Table 4 Coefficients (β) of the normalized (dimensionless) PLSR models of carbon fluxes (F_{CO_2}) for different deciduous forest study sites

Predictor variables	Bartlett Forest (NH)	Harvard Forest (MA)	UMBS Forest (MI)	Missouri Ozark (MO)	Silas Forest (NJ)	Willow Creek (WI)	Morgan Forest (IN)	Ohio Oak (OH)
RN	−0.20	−0.53	−0.19	−0.17	−0.22	−0.19	−0.23	−0.26
PAR	−0.17	−0.60	−0.15	−0.16	−0.15	−0.14	−0.18	−0.20
LHF	−0.36	−0.30	−0.54	−0.44	−0.58	−0.69	−0.64	−0.67
SHF	−0.08	0.68	0.07	0.11	0.11	0.22	0.24	0.23
VPD	0.14	0.11	0.09	0.04	−0.10	0.16	0.09	0.17
C_{CO_2}	0.18		0.26	0.28		0.15		0.13
TA	−0.05	0.01	0.07	0.06	0.07	0.00	−0.06	0.07
TS	−0.10	−0.06	0.03	0.02	0.10	0.00		0.00
WS	0.04	−0.06	0.11	0.07	0.06	0.02	0.04	0.10
UST	0.01	0.15	0.01	0.05	0.09	−0.04	0.02	−0.04
SWC	−0.03		−0.01	−0.03		0.05	−0.07	0.01
PLS components	3	5	3	3	4	4	3	4
R^2	0.64	0.64	0.69	0.55	0.63	0.75	0.74	0.81
RSR	0.60	0.59	0.56	0.67	0.61	0.50	0.51	0.44
MSE	0.37	0.35	0.31	0.45	0.37	0.25	0.26	0.19
β_{RHF}/β_T	2.5	8.3	5.0	6.3	4.1	4.5	5.8	4.2
β_{RHF}/β_W	11	6.8	5.4	5.9	6.0	18.8	16.6	7.2

Blank indicates missing data; RSR, the ratio of root-mean-square error to the standard deviation of observations, was calculated as: $RSR = \sqrt{\sum_{i=1}^{N}(F_{CO_2,i,mod} - F_{CO_2,i,obs})^2/N} \Big/ \sigma_{F_{CO_2},obs}$, where N is the total number of standardized observations of F_{CO_2}, $F_{CO_2,i,obs}$ and $F_{CO_2,i,mod}$ are the ith observed and predicted value of standardized F_{CO_2} (respectively), and $\sigma_{F_{CO_2},obs} = 1.0$ is the standard deviation of the observed, normalized F_{CO_2}; and MSE is the mean square error of predicted, normalized F_{CO_2}. The total relative linkages of the 'radiation-energy' (β_{RHF}), 'temperature-hydrology' (β_T), and 'aerodynamic' (β_W) components were computed, respectively, as: $\beta_{RHF} = \sqrt{\beta_{RN}^2 + \beta_{PAR}^2 + \beta_{LHF}^2 + \beta_{SHF}^2}$; $\beta_T = \sqrt{\beta_{TA}^2 + \beta_{TS}^2 + \beta_{VPD}^2 + \beta_{SWC}^2}$; and $\beta_W = \sqrt{\beta_{WS}^2 + \beta_{UST}^2}$

and 8 times stronger linkages with F_{CO_2} than that of the 'aerodynamic' and 'temperature-hydrology' groups, respectively. The relatively high coefficient of UST, compared to other sites, indicated a notable influence of canopy layer turbulent mixing on carbon flux transfer at this site.

UMBS

The VIP scores of LHF, PAR, RN, C_{CO_2}, VPD, and SHF were greater than (or equal to) 1.0, suggesting their strong relative linkages with F_{CO_2}; lower VIP scores indicated relatively weak to moderate linkages of F_{CO_2} with TA, TS, SWC, UST, and WS (Fig. 6c). Relative influence (β) of the PLSR model variables were mostly similar to that of VIPs (Table 4). LHF was the strongest predictor of F_{CO_2}; around 2–3.5 times stronger than that of the C_{CO_2}, RN and PAR. In contrast, the variables of WS, VPD, TA, and SHF showed around 5–7.5 times lower linkages with F_{CO_2}. Based on β, F_{CO_2} was very weakly linked with TS, UST, and SWC. The 'radiation-energy' component had around 5 times stronger linkages with F_{CO_2} than that of the 'temperature-hydrology' and 'aerodynamic' components.

Missouri Ozark Forest

The higher VIP scores of LHF, RN, PAR, C_{CO_2}, and VPD indicate that these variables were potentially strongly linked with F_{CO_2}; lower VIPs of the other variables refer to their low to moderate linkages with the carbon fluxes (Fig. 6d). These results were consistent with the regression modeling outcomes with the exceptions of SHF and VPD, which showed, respectively, relatively high and low coefficients with F_{CO_2}. The LHF was the strongest predictor of carbon fluxes, showing around 1.5 times stronger linkages than that of C_{CO_2}; 2.5–4 times stronger linkages than the PAR, RN, and SHF; and 6–8 times stronger linkages than TA, WS, and UST. The SWC and TS did not show noteworthy linkages with F_{CO_2}. Relatively lower model fitting efficiency and accuracy ($R^2 = 0.55$; RSR = 0.67; MSE = 0.45) also indicate the presence of more complicated (exceedingly nonlinear) carbon processes. Overall, the 'radiation-energy' group had approximately 6 times stronger linkages with F_{CO_2} than that of the 'aerodynamic' or 'temperature-hydrology' group (Table 4).

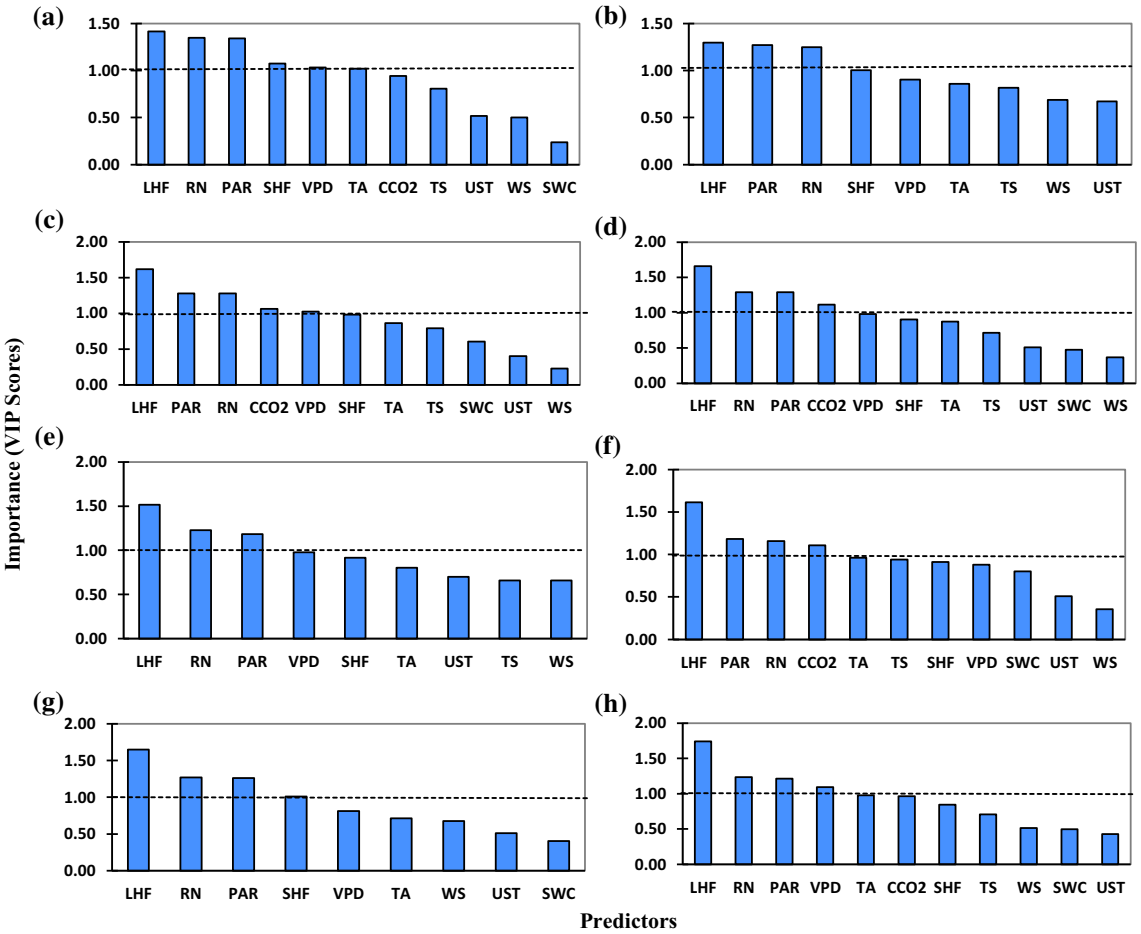

Fig. 6 Variable importance on the partial least squares projection (VIP) of different predictors for the response variable (F_{CO_2}) for (**a**) Bartlett Experimental Forest (NH), (**b**) Harvard Forest (MA), (**c**) UMBS Forest (MI), (**d**) Missouri Ozark Forest (MO), (**e**) Silas Little Forest (NJ), (**f**) Willow Creek Forest (WI), (**g**) Morgan Monroe State Forest (IN), and (**h**) Ohio Oak Openings Forest (OH). *Dashed line* indicates predictors having VIP score greater than unity (1.0)

Silas Little Experimental Forest

Based on the common outcomes of VIP scores (Fig. 6e) and PLSR coefficients (Table 4), the LHF, RN, and PAR had high linkages with F_{CO_2}, while the VPD, SHF, TA, TS, WS, and UST had relatively low to moderate carbon flux linkages. Per β, LHF was the strongest carbon flux predictor; around 2.5–4 times stronger than the RN and PAR; 5–6 times stronger than SHF, VPD, TS, and UST; and 8–10 times stronger than TA and WS. The ratio of β_{RHF}/β_T and β_{RHF}/β_W showed that the 'radiation-energy' component had around 4 and 6 times stronger linkages with F_{CO_2} than that of, respectively, the 'temperature-hydrology' and 'aerodynamic' components.

Willow Creek Forest

Both the VIP scores (Fig. 6f) and regression coefficients (Table 4) showed stronger linkages of carbon fluxes with the LHF, PAR, RN, C_{CO_2}, SHF, and VPD for this site; the soil moisture (SWC), wind speed (WS), and friction velocity (UST) had relatively low carbon flux linkages. The air and soil temperatures (TA and TS) showed negligible linkages with F_{CO_2}. Based on β, the predictive influence of LHF was around 3–4 times stronger than that of RN, PAR, SHF, C_{CO_2}; and 13–17 times stronger than SWC and UST. Overall, the 'radiation-energy' group had approximately 4.5 and 19 times stronger linkages with F_{CO_2} than that of the 'temperature-hydrology' and 'aerodynamic' groups, respectively (Table 4).

Morgan Monroe State Forest

The radiation and heat flux variables of LHF, RN, PAR, and SHF had high VIP scores (>1.0), indicating their potentially strong linkages with the carbon fluxes; small VIP scores indicated relatively weak linkages of F_{CO_2} with VPD, TA, WS, UST, and SWC. The PLSR model coefficients (β) (Table 4), as well as the outcomes of PCA

(Fig. 4c) and FA (Table 3), revealed nearly an identical linkage pattern. The LHF was the strongest predictor in regression, exhibiting around 2.5–3.5 times stronger linkages with F_{CO_2} than that of RN, PAR and SHF; 7–10 times stronger linkages than that of VPD, SWC, and TA; and around 16 times stronger linkages than WS. Based on β_{RHF}/β_T and β_{RHF}/β_W, the 'radiation-energy' component had around 6 and 16.5 times stronger linkages with the carbon fluxes than that of, respectively, the 'temperature-hydrology' and 'aerodynamic' components.

Ohio Oak Openings

The PLSR model had the highest fitting accuracy (MSE = 0.19; RSR = 0.44) and efficiency ($R^2 = 0.81$) for this site. Both the VIP scores (VIP >1; Fig. 6h) and regression coefficients (Table 4) showed relatively strong linkages of carbon fluxes with the LHF, PAR, RN, and VPD; the ambient carbon concentration (C_{CO_2}) had a moderate linkage, whereas the soil moisture (SWC) and friction velocity (UST) had very low carbon flux linkages. Although the sensible heat flux (SHF) and air temperature (TA) had moderate VIP scores, the associated PLSR model coefficients (β), as well as the PCA biplot (Fig. 4d) and FA results (Table 3), showed relatively strong influence of SHF and weak influence of TA in predicting F_{CO_2}. However, contrary to the VIP scores and the PCA and FA results, the wind speed (WS) had a moderately high regression coefficient (i.e., moderate carbon flux linkage), while the soil temperature (TS) showed a negligible coefficient (little linkage) with F_{CO_2}. Based on β, the predictive influence of LHF was around 2.5–4 times stronger than that of the RN, PAR, SHF, and VPD; 5–6.5 times stronger than C_{CO_2} and WS; and 9.5–17 times stronger than that of TA and UST. Overall, the 'radiation-energy' group had approximately 4 and 7 times stronger carbon flux linkages than that of the 'temperature-hydrology' and 'aerodynamic' group, respectively (Table 4).

Discussion

Linking Vertical Carbon Fluxes with the Climatic and Environmental Variables

For all eight deciduous forest sites, the 'radiation-energy' component of RN, PAR, SHF, and LHF was strongly linked with the canopy-level CO_2 fluxes. Previous research (e.g., Jung et al. 2011; Schmidt et al. 2011; Morales et al. 2005; Zhang et al. 2005; Baker et al. 2003; Sellers et al. 1997) also reported a similar finding. However, most studies considered the latent heat flux (LHF) and sensible heat flux (SHF), along with F_{CO_2}, as the response variables as functions of some common drivers such as the radiation, temperature, vapor pressure, etc. In contrast, alike Melesse and Hanley (2005), our study included the two heat fluxes in the matrix of predictor variables; quantifying their relatively high linkages with F_{CO_2} exchanges within a large set of climatic and environmental variables.

The mutual correlations among the three fluxes of F_{CO_2}, SHF and LHF could partly stem from the common dynamic term (vertical wind speed fluctuations) of their eddy covariance measurement equations (Launiainen et al. 2005); however, their interrelationships identified in this study were mostly process-oriented. The carbon and energy fluxes together represent the ecosystem's biological exchanges with atmosphere (Baldocchi et al. 2001). Heat fluxes (LHF and SHF) help to maintain balance in atmospheric radiation through evapotranspiration (ET) and turbulent energy diffusion (Sellers et al. 1997). Plants' stomata tend to close with increasing transpiration and LHF slows down to maintain ecosystem water budget, indirectly affecting carbon flux exchanges (Heber et al. 1986). Furthermore, the relative weight (i.e., ratio) between the canopy-level photosynthesis and ET is a moderate function of atmospheric humidity deficit (Baldocchi and Meyers 1998), indicating ET-control on high rate of canopy growth.

Although the ambient CO_2 concentration (C_{CO_2}) was not available for three of the eight study sites, it showed relatively moderate to strong linkages with carbon fluxes (F_{CO_2}) for all available sites. Elevated C_{CO_2} stimulates the photosynthesis by increasing the carboxylation and oxygenation, leading to a fast plant growth and ultimately increasing the litter production and soil carbon storage (Masle 2000). Furthermore, high atmospheric carbon concentrations often facilitate more efficient use of available soil water for plant growth and productivity (Schlesinger 1999). In contrast, Drake et al. (1999) reported a negative functional relation between plant respiration and C_{CO_2}.

The 'temperature-hydrology' component (formed by TA, TS, VPD, and SWC) had a moderate linkage with the vertical carbon fluxes for almost all study forests. Water availability, temperature, and light can play a significant role in plants' energy and water exchanges by transforming stomatal aperture (short-term) and density (long-term) (Haworth et al. 2011). High vapor pressure deficit (VPD) can lead to stomatal closure (Loescher et al. 2003), affecting the plant-atmospheric energy (latent heat) and carbon fluxes. VPD also loaded appreciably in the 'radiation-energy' factor (Table 3), which is consistent with previous studies (e.g., Lund et al. 2010). Although soil moisture (SWC) at most study sites grouped with the

velocity variables in the PCA biplots (Figs. 2, 3), more detailed (3-dimensional) information emerged in FA analysis showing its appropriate association with the 'temperature-hydrology' component (Table 3). The SWC can contribute to the plants' photosynthesis by influencing the water potential difference between the tree leaves and root system. However, our five layer analysis (Pearson correlation, PCA, FA, PLSR-VIP, and PLSR-BETA) with half-hourly data of different years from different US deciduous forests indicated a relatively low predictive influence of soil moisture on vertical CO_2 fluxes for the small time-scale. The 'aerodynamic' component (WS, UST) was also relatively weakly linked with the carbon flux exchanges. Mechanistically UST and WS are not direct contributors of carbon fluxes (Wilson et al. 2002); rather they influence the boundary layer vertical mixing to facilitate transport. Future studies should focus into the attributions of these low carbon flux linkages of SWC, WS, and UST; for example, whether aggregating data in larger time-scales (e.g., daily, weekly, monthly, yearly) reveal a more notable linkage of SWC with the F_{CO_2}.

Despite the gradients of a variable canopy height (9.52–27 m), different climatic regimes (humid to temperate to nearly extreme), and diverse soil morphology among the study sites, the relative linkages of major process components, as well as individual climate and environmental variables, with F_{CO_2} were essentially similar. Our findings of the relative carbon flux linkages, using a simple method on half-hourly data for exclusively deciduous forests, complement the findings of Schmidt et al. (2011) that applied a complex neural network approach to analyze daily averaged data for different types of ecosystems, including five deciduous forests. For example, both studies conclude that air temperature had a much lower effect on turbulent carbon fluxes than that of other variables such as the radiations; the same is true for the relatively weak carbon flux linkage of wind and friction velocities. However, Schmidt et al.(2011) reported notably strong linkages of daily precipitation and soil temperature with F_{CO_2}, whereas our study found relatively low to moderate carbon flux linkages of soil moisture and soil temperature for the smaller (half-hourly) time-scale.

Data Quality and Uncertainty

Error and uncertainty associated with the eddy fluxes and other variables can cause superfluous biases in analysis and modeling (Williams et al. 2009). Our two-step data filtering procedure substantially eliminated the unsuitable data representing gaps (for more than two variables in one half-hourly panel) and unreasonable spikes (outliers) from the final data sets, which were analyzed to derive and report the results. Furthermore, previous research (Schmidt et al. 2012) reported relatively low measurement uncertainties for canopy-level meteorological variables (relative error $\leq$2 %), heat fluxes (relative error = 1.7–5.2 %) and vertical CO_2 fluxes (F_{CO_2}) (relative error $\leq$8.2 %) of the AmeriFLUX network. Schmidt et al. (2012) also suggested high quality research applicability of the AmeriFLUX data. Nevertheless, it is possible that our analysis and modeling with half-hourly data may not be completely free from the effect of random sampling errors of measurements (Baldocchi 2003).

Explanatory Modeling and Analysis

Theoretically process-based carbon dynamics models should be more reliable than the empirical models, but all the relevant processes are not understood yet (Keenan et al. 2011). This reemphasizes the importance of developing data-informed carbon flux modeling system (Keenan et al. 2012), which requires proper mechanistic judgments in the selection of model variables. The multicollinearity effect generally provides biased models with the traditional least squares regression method, which may be partially resolved by eliminating predictors stepwise (backward or forward) or simultaneously based on statistical significance. However, this elimination sometimes results in removal of variables that has important mechanistic basis, hampering the evaluation of comparative linkages of relevant predictors with the response variable. Our dimensionless, optimal PLSR modeling approach (capturing maximum system variance and mechanisms) provided the flexibility and statistical stability for retaining all predictors since regression was primarily done with the orthogonal PLS components and then transformed to the original domain. Predictions with the linear PLSR models were quite impressive (R^2 = 0.55–0.81), as compared to the nonlinear half-hourly data-driven models of Byrne et al. (2005) for GPP (R^2 = 0.78–0.81) and respiration (R^2 = 0.86–0.83), hourly mechanistic model of Wu and Chen (2013) for deciduous forest carbon fluxes (R^2 = 0.66–0.91), for example.

The PLSR models could not satisfactorily predict extremely positive (upward) and negative (downward) carbon fluxes likely due to the linear structure. Primary production (GPP) and respiration have different mechanisms and are not necessarily controlled by the same set of drivers. Representing these two distinct processes by one of set of regression coefficients at the half-hourly scale can contribute errors to the model predictions. Seasonal variability in terrestrial carbon fluxes can also cause uncertainty in a data-driven model developed with half-hourly data (Jarvis et al. 1997; Xu and Baldocchi 2004). Furthermore, instead of including multiple years of observations for each study site, single-year data, encompassing a

5-year time frame (2006–2011) among the eight deciduous study forests, were chosen for our analysis and modeling. Although a 10-fold cross-validation improved consistency in model fitting and robustness of estimated parameters, incorporation of single (rather than multiple) year data for an individual site could be seen as a limitation of this study.

Carbon fluxes can strongly respond to precipitation and vegetation productivity (Piao et al. 2013; Schmidt et al. 2011) considering their control over the long-term carbon balance. We used soil moisture (SWC) data as a surrogate for precipitation subject to the lack of availability of half-hourly precipitations. Since our study goal was to determine the relative linkages of mainly climate and environmental variables with the canopy-level vertical carbon fluxes, biological variables such as the canopy leaf-area-index (LAI) was excluded from the predictor data matrix. Further, variables such as SWC and LAI are less likely to change much over a half-hourly interval; our results showed little carbon flux linkages of SWC for the half-hour scale at the US deciduous forests. Although ambient atmospheric concentrations of CO_2 were included in our data matrices, we did not explicitly incorporate anthropogenic carbon sources, which could influence the large-scale ecosystem carbon emissions. Exclusion of these process components, apart from the linear structure of PLSR, could have contributed to the reduction of our model fitting accuracy (MSE; RSR) and efficiency (R^2) for different study sites.

Building linkages between statistical and mechanistic modeling approaches has been a major challenge in ecological modeling research (Larocque et al. 2011). Issues such as the data requirements, complex parameterizations, prediction uncertainties, computational expenses, and expert knowledge basis of available mechanistic carbon-cycle models highlight the importance of developing relatively simple models without conceding the representation of important processes at the relevant spatiotemporal scales. Our study presented a simple, systematic multivariate approach to identify the dominant process components by classifying the relevant climate and environmental variables, quantifying their relative linkages with the canopy-level vertical carbon fluxes. The modeling and analysis provides an objective, empirical foundation to obtain crucial mechanistic insights a priori; complementing process-based model building with a warranted complexity.

Conclusions

We used a data-analytics method to determine the relative linkages of different climate and environmental variables with the canopy-level, half-hourly CO_2 fluxes of US deciduous forests. Three biophysical process components were identified to adequately explain the canopy-level, vertical CO_2 fluxes. The 'radiation-energy' component had the strongest linkages with the canopy-level CO_2 fluxes. The 'temperature-hydrology' component showed low to moderate carbon flux linkages. The 'aerodynamic' component was relatively weakly connected with the carbon fluxes. The relative linkage of ambient CO_2 concentrations with the vertical carbon fluxes was moderate to strong among different sites. The latent heat flux was the most influential predictor of instantaneous CO_2 fluxes at all study sites except for Harvard Forest. On average, the 'radiation-energy' component showed around 5 and 8 times stronger carbon flux linkages than that of the 'temperature-hydrology' and 'aerodynamic' components, respectively. The similarity of observed patterns among different study sites (representing sharp gradients in canopy heights, climatic regimes, and soil formations) indicates that the findings are potentially transferrable to other deciduous forests around the world. The observed similarities also highlight the scope of developing robust, parsimonious models for appropriate predictions of ecosystem carbon fluxes and potential sequestrations under a changing climate and environment. Relatively good model accuracy and efficiency reiterate the usefulness of multivariate analytics models for gap-filling in time-series of instantaneous flux data.

Future research should investigate the relative linkage patterns by aggregating data in larger time-scales (e.g., daily, weekly, monthly). Multi-scale linkage patterns in other terrestrial ecosystems should also be investigated and compared. More advanced data-analytics approaches such as the system and network modeling, machine learning, and fuzzy logic may also contribute toward developing a robust understanding and prediction of ecosystem carbon fluxes at different spatiotemporal scales.

Acknowledgments The research was funded by grants from the National Science Foundation's Environmental Sustainability Program (NSF CBET Award No. 1336911), and from the National Oceanic and Atmospheric Administration (NOAA NERRA Grant No. NA09NOS4190153). The supports are gratefully acknowledged. We also acknowledge the availability and usefulness of AmeriFLUX database, as funded by the US Department of Energy's Office of Science. The statements, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the views of NSF or NOAA. Thanks to the reviewers and the Editor for providing insightful comments.

References

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Automat Contr 19(6):716–723

Aubinet M, Grelle A, Ibrom A et al (1999) Estimates of the annual net carbon and water exchange of forests: the EUROFLUX methodology. Adv Ecol Res 30:113–175

Baker I, Denning AS, Hanan N et al (2003) Simulated and observed fluxes of sensible and latent heat and CO2 at the WLEF-TV tower using SiB2.5. Glob Chang Biol 9(9):1262–1277

Baldocchi DD (2003) Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future. Glob Chang Biol 9(4):479–492

Baldocchi DD, Meyers T (1998) On using eco-physiological, micrometeorological and biogeochemical theory to evaluate carbon dioxide, water vapor and trace gas fluxes over vegetation: a perspective. Agric For Meteorol 90(1–2):1–25

Baldocchi DD, Falge E, Gu L et al (2001) FLUXNET: a new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. Bull Am Meteorol Soc 82(11):2415–2434

Barr AG, Black TA, Hogg EH, 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 For Meteorol 126(3):237–255

Beer C, Reichstein M, Tomelleri E et al (2010) Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate. Science 329:834–838

Braswell BH, Sacks WJ, Linder E, Schimel DS (2005) Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations. Glob Chang Biol 11(2):335–355

Byrne KA, Kiely G, Leahy P (2005) CO_2 fluxes in adjacent new and permanent temperate grasslands. Agric For Meteorol 135(1):82–92

Cao M, Woodward FI (1998) Dynamic responses of terrestrial ecosystem carbon cycling to global climate change. Nature 393(6682):249–252

Carrara A, Kowalski AS, Neirynck J, Janssens IA, Yuste JC, Ceulemans R (2003) Net ecosystem CO_2 exchange of mixed forest in Belgium over 5 years. Agric For Meteorol 119(3):209–227

Carvalhais N, Seixas J, Myneni R (2005) Modeling net ecosystem productivity scale issues and regional application to the Iberian Peninsula. In: Seventh international carbon dioxide conference (ICDC7) September 25–30

Chen M, Zhuang Q, Cook DR et al (2011) Quantification of terrestrial ecosystem carbon dynamics in the conterminous United States combining a process-based biogeochemical model and MODIS and AmeriFlux data. Biogeosci Discuss 8(2):2721–2773

Chong IG, Jun CH (2005) Performance of some variable selection methods when multicollinearity is present. Chemometr Intell Lab Syst 78(1):103–112

de Jong S (1993) SIMPLS: an alternative approach to partial least squares regression. Chemometr Intell Lab Syst 18(3):251–263

Desai AR (2010) Climatic and phenological controls on coherent regional interannual variability of carbon dioxide flux in a heterogeneous landscape. J Geophys Res. doi:10.1029/2010JG001423

Dragon K (2006) Application of factor analysis to study contamination of a semi-confined aquifer (Wielkopolska Buried Valley aquifer, Poland). J Hydrol 331(1):272–279

Drake BG, Azcon-Bieto J, Berry J et al (1999) Does elevated atmospheric CO_2 concentration inhibit mitochondrial respiration in green plants? Plant Cell Environ 22(6):649–657

Falge E, Baldocchi D, Olson R et al (2001) Gap filling strategies for long term energy flux data sets. Agric For Meteorol 107(1):71–77

Fuentes JD, Wang D (1999) On the seasonality of isoprene emissions from a mixed temperate forest. Ecol Appl 9(4):1118–1131

Geider RJ, Delucia EH, Falkowski PG et al (2001) Primary productivity of planet earth: biological determinants and physical constraints in terrestrial and aquatic habitats. Glob Chang Biol 7(8):849–882

Gilmanov TG, Parton WJ, Ojima DS (1997) Testing the 'CENTURY' ecosystem level model on data sets from eight grassland sites in the former USSR representing a wide climatic/soil gradient. Ecol Modell 96(1):191–210

Gove JH, Holinger DY (2006) Application of a dual unscented Kalman filter for simultaneous state and parameter estimation in problems of surface-atmospheric exchange. J Geophys Res Atmos 111:1–21

Grant RF, Baldocchi DD, Ma S (2012) Ecological controls on net ecosystem productivity of a seasonally dry annual grassland under current and future climates: modelling with ecosys. Agric For Meteorol 152:189–200

Hargrove WW, Hoffman FM (2005) Potential of multivariate quantitative methods for delineation and visualization of ecoregions. Environ Manage 34(1):S39–S60

Haworth M, Elliott-Kingston C, McElwain JC (2011) Stomatal control as a driver of plant evolution. J Exp Bot 62(8):2419–2423

Heber U, Neimanis S, Lange OL (1986) Stomatal aperture, photosynthesis and water fluxes in mesophyll cells as affected by the abscission of leaves. Simultaneous measurements of gas exchange, light scattering and chlorphyll fluorescence. Planta 167(4):554–562

Heimann M, Reichstein M (2008) Terrestrial ecosystem carbon dynamics and climate feedbacks. Nature 451(7176):289–292

Hollinger DY, Aber J, Dail B et al (2004) Spatial and temporal variability in forest–atmosphere CO_2 exchange. Glob Chang Biol 10(10):1689–1706

Hubert M, Branden KV (2003) Robust methods for partial least squares regression. J Chemom 17(10):537–549

Hui D, Wan S, Su B, Katul G, Monson R, Luo Y (2004) Gap-filling missing data in eddy covariance measurements using multiple imputation (MI) for annual estimations. Agric For Meteorol 121(1):93–111

Jahan N, Gan TY (2013) Developing a gross primary production model for coniferous forests of northeastern USA from MODIS data. Int J Appl Earth Obs Geoinf 25:11–20

Jarvis PG, Massheder JM, Hale SE, Moncrieff JB, Rayment M, Scott SL (1997) Seasonal variation of carbon dioxide, water vapor, and energy exchanges of a boreal black spruce forest. J Geophys Res Atmos (1984–2012) 102(D24):28953–28966

Jolliffe IT (1993) Principal component analysis: a beginner's guide-II. Pitfalls, myths and extensions. Weather 48(8):246–253

Jung M, Reichstein M, Margolis HA et al (2011) Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. J Geophys Res. doi:10.1029/2010JG001566

Keenan T, Maria Serra J, Lloret F, Ninyerola M, Sabate S (2011) Predicting the future of forests in the Mediterranean under climate change, with niche-and process-based models: CO_2 matters! Glob Chang Biol 17(1):565–579

Keenan TF, Davidson E, Moffat AM, Munger W, Richardson AD (2012) Using model-data fusion to interpret past trends, and quantify uncertainties in future projections, of terrestrial ecosystem carbon cycling. Glob Chang Biol 18(8):2555–2569

Kuhn M, Johnson K (2013) Applied predictive modeling. Springer, New York

Larocque GR, Mailly D, Yue TX et al (2011) Common challenges for ecological modelling: synthesis of facilitated discussions held at the symposia organized for the 2009 conference of the International Society for Ecological Modelling in Quebec City, Canada, (October 6–9, 2009). Ecol Modell 222(14):2456–2468

Launiainen S, Rinne J, Pumpanen J et al (2005) Eddy covariance measurements of CO_2 and sensible and latent fluxes during a full year in a boreal pine forest trunk-space. Boreal Environ Res 10(6):569–588

Li Z, Yu G, Xiao X et al (2007) Modeling gross primary production of alpine ecosystems in the Tibetan Plateau using MODIS images and climate data. Remote Sens Environ 107(3):510–519

Liu CW, Lin KH, Kuo YM (2003) Science Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. Sci Total Environ 313(1):77–89

Lloyd J, Taylor JA (1994) On the temperature dependence of soil respiration. Funct Ecol 8:315–323

Loescher HW, Oberbauer SF, Gholz HL, Clark DB (2003) Environmental controls on net ecosystem-level carbon exchange and productivity in a Central American tropical wet forest. Glob Chang Biol 9(3):396–412

Lund M, Lafleur PM, Roulet NT et al (2010) Variability in exchange of CO_2 across 12 northern peatland and tundra sites. Glob Chang Biol 16(9):2436–2448

Mahbub P, Ayoko GA, Goonetilleke A, Egodawatta P, Kokot S (2010) Impacts of traffic and rainfall characteristics on heavy metals build-up and wash-off from urban roads. Environ Sci Technol 44(23):8904–8910

Mäkelä A, Pulkkinen M, Kolari P et al (2008) Developing an empirical model of stand GPP with the LUE approach: analysis of eddy covariance data at five contrasting conifer sites in Europe. Glob Chang Biol 14(1):92–108

Masle J (2000) The effects of elevated CO_2 concentrations on cell division rates, growth patterns, and blade anatomy in young wheat plants are modulated by factors related to leaf position, vernalization, and genotype. Plant Physiol 122(4):1399–1416

Melesse AM, Hanley RS (2005) Artificial neural network application for multi-ecosystem carbon flux simulation. Ecol Modell 189(3):305–314

Moorcroft PR, Hurtt GC, Pacala SW (2001) A method for scaling vegetation dynamics: the ecosystem demography model (ED). Ecol Monogr 71(4):557–586

Morales P, Sykes MT, Prentice IC (2005) Comparing and evaluating process-based ecosystem model predictions of carbon and water fluxes in major European forest biomes. Glob Chang Biol 11(12):2211–2233

Moriasi DN, Arnold JG, Van Liew MW, Bingner RL, Harmel RD, Veith TL (2007) Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Am Soc Agric Biol Eng 50(3):885–900

Oechel WC, Vourlitis GL, Verfaillie J et al (2000) A scaling approach for quantifying the net CO_2 flux of the Kuparuk River Basin, Alaska. Glob Chang Biol 6(S1):160–173

Panda UC, Sundaray SK, Rath P, Nayak BB, Bhatta D (2006) Application of factor and cluster analysis for characterization of river and estuarine water systems—a case study: Mahanadi River (India). J Hydrol 331(3):434–445

Peres-Neto PR, Jackson DA, Somers KM (2003) Giving meaningful interpretation to ordination axes: assessing loading significance in principal component analysis. Ecology 84(9):2347–2363

Piao S, Sitch S, Ciais P et al (2013) Evaluation of terrestrial carbon cycle models for their response to climate variability and to CO_2 trends. Glob Chang Biol 19:2117–2132

Post WM, Pastor J (1996) Linkages—an individual-based forest ecosystem model. Clim Change 34(2):253–261

Richardson AD, Braswell BH, Hollinger DY et al (2006) Comparing simple respiration models for eddy flux and dynamic chamber data. Agric For Meteorol 141(2):219–234

Running SW, Coughlan JC (1988) A general model of forest ecosystem processes for regional applications I. Hydrologic balance, canopy gas exchange and primary production processes. Ecol Modell 42(2):125–154

Running SW, Gower ST (1991) FOREST-BGC, a general model of forest ecosystem processes for regional applications. II. Dynamic carbon allocation and nitrogen budgets. Tree Physiol 9(1–2):147–160

Schimel DS, House JI, Hibbard KA et al (2001) Recent patterns and mechanisms of carbon exchange by terrestrial ecosystems. Nature 414(6860):169–172

Schlesinger WH (1999) Carbon sequestration in soils. Science 284(5423):2095

Schmidt A, Hanson C, Kathilankal J, Law BE (2011) Classification and assessment of turbulent fluxes above ecosystems in North-America with self-organizing feature map networks. Agric For Meteorol 151(4):508–520

Schmidt A, Hanson C, Chan WS, Law BE (2012) Empirical assessment of uncertainties of meteorological parameters and turbulent fluxes in the AmeriFlux network. J Geophys Res. doi:10.1029/2012JG002100

Schubert P, Lagergren F, Aurela M et al (2012) Modeling GPP in the Nordic forest landscape with MODIS time series data—comparison with the MODIS GPP product. Remote Sens Environ 126:136–147

Sellers PJ, Dickinson RE, Randall DA et al (1997) Modeling the exchanges of energy, water, and carbon between continents and the atmosphere. Science 275(5299):502–509

Shir CC, Bornstein RD (1977) Eddy exchange coefficients in numerical models of the planetary boundary layer. Bound Layer Meteorol 11(2):171–185

Sims DA, Rahman AF, Cordova VD et al (2008) A new model of gross primary productivity for North American ecosystems based solely on the enhanced vegetation index and land surface temperature from MODIS. Remote Sens Environ 112(4):1633–1646

Stauch VJ, Jarvis AJ (2006) A semi-parametric gap-filling model for eddy covariance CO_2 flux time series data. Glob Chang Biol 12(9):1707–1716

Turner DP, Ritts WD, Styles JM, Yang Z, Cohen WB, Law BE, Thornton PE (2006) A diagnostic carbon flux model to monitor the effects of disturbance and interannual variation in climate on regional NEP. Tellus B 58(5):476–490

White A, Cannell MG, Friend AD (1999) Climate change impacts on ecosystems and the terrestrial carbon sink: a new assessment. Glob Environ Change 9:S21–S30

Williams M, Richardson AD, Reichstein M et al (2009) Improving land surface models with FLUXNET data. Biogeosci Discuss 6(2):2785–2835

Wilson K, Goldstein A, Falge E et al (2002) Energy balance closure at FLUXNET sites. Agric For Meteorol 113(1):223–243

Wold H (1966) Estimation of principal components and related models by iterative least squares. In: Krishnaiah PR (ed) Multivaraite analysis. Academic Press, New York, pp 391–420

Wold H (1982) Soft modelling: the basic design and some extensions. In: Jöreskog KG, Wold H (eds) Systems under indirect observation, part II. North Holland Press, Amsterdam

Wold S, Johansson M, Cocchi M (1993) PLS-partial least squares projections to latent structures. In: Kubinyi H (ed) 3D QSAR in drug design, vol 1. Kluwer Academic Publishers, Netherlands, pp 523–550

Wold S, Sjöström M, Eriksson L (2001) PLS-regression: a basic tool of chemometrics. Chemometr Intell Lab Syst 58(2):109–130

Wu C, Chen JM (2013) Deriving a new phenological indicator of interannual net carbon exchange in contrasting boreal deciduous and evergreen forests. Ecol Indic 24:113–119

Wylie BK, Fosnight EA, Gilmanov TG, Frank AB, Morgan JA, Haferkamp MR, Meyers TP (2007) Adaptive data-driven models for estimating carbon fluxes in the Northern Great Plains. Remote Sens Environ 106(4):399–413

Xu L, Baldocchi DD (2004) Seasonal variation in carbon dioxide exchange over a Mediterranean annual grassland in California. Agric For Meteorol 123(1):79–96

Zhang Y, Grant RF, Flanagan LB, Wang S, Verseghy DL (2005) Modelling CO_2 and energy exchanges in a northern semiarid grassland using the carbon-and nitrogen-coupled Canadian Land Surface Scheme (C-CLASS). Ecol Modell 181(4):591–596