Revisiting and attributing the global controls on terrestrial ecosystem functions of climate and plant traits at FLUXNET sites ~~with~~via causal ~~networks~~graphical models

Haiyang Shi[1,2,4,56], Geping ~~Luo~~[1,2]Luo[2,3,54,6], Olaf ~~Hellwich~~[6]Hellwich[7], Alishir ~~Kurban~~[1,2]Kurban[2,3,54,6], Philippe De ~~Maeyer~~[1,2,4]Maeyer[2,3,5,6] and Tim Van de ~~Voorde~~[4,5]Voorde[5,6]

[11] School of Earth Sciences and Engineering, Hohai University, Nanjing 211100, China.
[2] State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi, Xinjiang, 830011, China.
[23] College of Resources and Environment, University of the Chinese Academy of Sciences, 19 (A) Yuquan Road, Beijing, 100049, China.
[34] Research Centre for Ecology and Environment of Central Asia, Chinese Academy of Sciences, Urumqi, China.
[45] Department of Geography, Ghent University, Ghent 9000, Belgium.
[56] Sino-Belgian Joint Laboratory of Geo-Information, Ghent, Belgium.
[67] Department of Computer Vision & Remote Sensing, Technische Universität Berlin, 10587 Berlin, Germany.

Correspondence to: Geping Luo (luogp@ms.xjb.ac.cn) and Olaf Hellwich (olaf.hellwich@tu-berlin.de)

Submitted to: *Biogeosciences*

Abstract

Using statistical methods that ~~do~~ not ~~emphasize~~directly representing the ~~systematic~~ causality between variables to attribute climate and plant traits to control ecosystem function may produce biased perceptions. We revisit this issue using a causal graphical model, Bayesian network (BN~~)~~), capable of quantifying causality by conditional probability tables. Based on expert knowledge and climate, vegetation, and ecosystem function data from the FLUXNET flux stations, we constructed a BN containing the causal relationship of 'climate-plant trait-ecosystem function'. Based on the sensitivity analysis function of the BN, we attributed the ~~control~~controls of climate and plant traits to ecosystem function and compared the results with those based on Random forests and correlation analysis. The main conclusions of this study include: BN can be used for the quantification of causal relationships between complex ecosystems ~~and climatic and environmental systems,~~in response to climate change and enables the analysis of indirect effects among variables. ~~The control of ecosystem function~~Compared to BN, the feature importance difference between 'VPD and CSWI' and 'LAImax and Hc' reported by ~~climate~~ Random forests is higher and can be overestimated. With the causality relation between correlated variables ~~(especially mean temperature and mean vapor pressure deficit) may have been underestimated previously, and~~constructed, BN-based sensitivity analysis can reduce the uncertainty in quantifying the importance of correlated variables. The understanding of the mechanism of indirect effects of climate variables on ecosystem function through plant traits ~~should be emphasized in future studies. Further inclusion of temporal information in BN holds promise for improving the analysis of lagged effects and interactions and feedback effects between variables~~can be deepened by the chain casuality quantification in BNs.

1 Introduction

~~Terrestrial ecosystems provide a variety of important ecosystem functions for our society (Manning et al., 2018). It is essential to understand the potential changes in ecosystem functions in the context of global climate change (Grimm et al., 2013). The response of terrestrial ecosystem function to changes in climate change, plant traits, and environmental conditions, and the corresponding mechanisms, are complex due to enormous spatial and temporal variations across ecosystems, climate zones, and also space-time scales (Diaz and Cabido, 1997; Madani et al., 2018; Myers-Smith et al., 2019). Given the enormous variations, on the global scale, these issues have not been clarified well.~~

~~In the past decades, measurements of ecosystem functions are increasingly available to support studies of the relations between ecosystem functions and climate and environmental systems. For example, eddy-covariance flux tower observations (Baldocchi, 2014) for carbon flux (i.e., net ecosystem exchange (NEE)) and water flux (i.e., evapotranspiration (ET)) have been widely used to investigate changes in ecosystem functions and their responses to climate change, vegetation condition changes, etc (Jung et al., 2020, 2010; Migliavacca et al., 2021; Peaucelle et al., 2019). With the increase in such observations, various statistical analysis methods such as emerging machine learning (Barnes et al., 2021; Migliavacca et al., 2021; Reichstein et al., 2019; Shi et al., 2022b, a, 2020b; Tramontana et al., 2016) have been used to mine the hidden information on the effects of climate change and its induced changes in vegetation, etc. on ecosystem function variables such as carbon and~~

~~water flux, which has not been understood in depth by process-based models (e.g., biogeochemistry models (Sakschewski et al., 2016)). For example, using Random forests (RF) and principal component analysis (PCA), a recent study (Migliavacca et al., 2021) quantified the three main axes of terrestrial ecosystem function and their drivers based on observations of carbon and water fluxes of FLUXNET (Pastorello et al., 2020) and various climate and plant trait variables. Generally, data-driven approaches have become increasingly important recently in this area (Reichstein et al., 2019).~~

~~However, compared to the process-based models, most of these data-driven approaches lack representation of the systematic causality and detailed processes in the relations between ecosystem function and climate and environments, despite the widely recognized complex causal interactions of ecosystems with climate and environmental systems (Reichstein et al., 2014). Conventional methods such as multiple linear regression have been questioned in attribution studies of the relationship between climate and the carbon cycle (Wang et al., 2022). For example, the use of multiple linear regression may underestimate the direct effect of soil moisture possibly due to the covariance between variables (Wang et al., 2022). For machine learning techniques, although current common algorithms such as RF (Migliavacca et al., 2021) can report the importance of features (IMP) to measure their contributions to the prediction model, IMP-based attribution to the target variable can also be unreliable when we aim to explain systematic causality (Gregorutti et al., 2017). Therefore, it is commonly important to recognize the difference between correlation and causality in these approaches and emphasize the systematic causality in the systems and also detailed causal relations between features in a data-driven approach.~~

~~Bayesian network (BN) is a causal model based on conditional probability representation (Friedman et al., 1997; Pearl, 1985) that characterizes the transmission of cause and effect through conditional probabilities between variables. Currently, BN has been used in modeling causal relationships in many fields and has demonstrated advantages in causal interpretation, including in the fields such as hydrology and ecology (Chan et al., 2010; Keshtkar et al., 2013; Milns et al., 2010; Pollino et al., 2007; Shi et al., 2021a, b; Trifonova et al., 2015). However, BN has rarely been used in the study of attribution of changes in ecosystem function. Therefore, this study used BN to attribute the controls of climate and plant traits on ecosystem function by quantifying the causal relationships involved. The data used are from a previous study (Migliavacca et al., 2021) which extracted ecosystem function, climate, and plant trait variables for FLUXNET flux stations. The construction of the causal structure of BN referred to the previous expert knowledge of this system (Reichstein et al., 2014). Further, by comparing BN-based attribution analysis, linear correlation analysis, and RF-based IMP reported by the previous study (Migliavacca et al., 2021), we investigated the adding-values of using BN for causal analysis and discussed its prospects in this paper.~~

~~2 Methodology~~

~~2.1 Data~~

~~The used variables (Table 1) include the carbon and water fluxes of the FLUXNET flux tower sites and the ecosystem function variables derived from them, and information on the corresponding climatic~~Ecosystem function is the capacity of natural processes and components to provide goods and services that satisfy human needs, either directly or indirectly (de Groot et al., 2002). Ecosystem functions include the physicochemical and biological processes within the ecosystem to maintain terrestrial life. Terrestrial ecosystems have provided a variety of important ecosystem functions for our society (Manning et al., 2018). Plant traits' role as important determinants of ecosystem functions has been widely recognized (Chapin Iii et al., 2000), and various trait syndromes can result in distinct broad differences in ecosystem functions (Reichstein et al., 2014). In the context of global climate change, it is also essential to understand the potential changes in ecosystem functions (Grimm et al., 2013). The response of terrestrial ecosystem function to changes in climate, plant traits, and the corresponding mechanisms, are complex due to enormous spatial and temporal variations across ecosystems, climate zones, and also space-time scales (Diaz and Cabido, 1997; Madani et al., 2018; Myers-Smith et al., 2019). Given the enormous variations, on the global scale, these issues have not been clarified well.

In the past decades, measurements of ecosystem functions are increasingly available to support studies of the relations between ecosystem functions and climate variables. For example, eddy-covariance flux tower observations (Baldocchi, 2014) for carbon flux (i.e., net ecosystem exchange (NEE)) and water flux (i.e., evapotranspiration (ET)) have been widely used to investigate changes in ecosystem functions and their responses to climate change, vegetation condition changes, etc (Jung et al., 2020, 2010; Migliavacca et al., 2021; Peaucelle et al., 2019). With the increase in such observations, various statistical analysis methods such as emerging machine learning (Barnes et al., 2021; Migliavacca et al., 2021; Reichstein et al., 2019; Shi et al., 2022b, a, 2020b; Tramontana et al., 2016) have been used to mine the hidden information on the effects of climate change and its induced changes in vegetation, etc. on ecosystem function variables such as carbon and water flux, which has not been understood in depth by process-based models (e.g., biogeochemistry models (Sakschewski et al., 2016)). For example, using Random forests (RF) and principal component analysis (PCA), a recent study (Migliavacca et al., 2021) quantified the three main axes of terrestrial ecosystem function and their drivers based on observations of carbon and water fluxes of FLUXNET (Pastorello et al., 2020) and various climate and plant trait variables. Generally, data-driven approaches have become increasingly important recently in this area (Reichstein et al., 2019).

However, compared to the process-based models, most of these data-driven approaches lack representation of the causality and detailed processes in the relations between ecosystem function and climate, despite the widely recognized complex causal interactions of ecosystems with climate systems (Reichstein et al., 2014). Conventional methods such as multiple linear regression have been questioned in attribution studies of the relationship between climate and the carbon cycle (Wang et al., 2022). For example, the use of multiple linear regression may underestimate the direct effect of soil moisture possibly due to the covariance between variables (Wang et al., 2022). For machine learning techniques, current common algorithms such as RF (Migliavacca et al., 2021) can report the importance of features (IMP) to measure their contributions to the prediction model. However, IMP-based attribution to the target variable can also be unreliable if considerable confounders and correlations between predictor variables exist (Strobl et al., 2008; Toloşi and Lengauer, 2011). The less relevant predictors can replace the predictive predictors (due to correlation) and thus receive undeserved high feature importance (Strobl et al., 2008). Correlations between predictors can lead to biased feature-importance-based findings. It is thus important to recognize the difference between correlation and causality in these approaches, represent detailed causal relations between features, rather than the unreliable feature importance rankings generated from correlated features.

Bayesian network (BN) is a causal graphical model based on conditional probability representation (Friedman et al., 1997; Pearl, 1985) that characterizes the transmission of cause and effect through conditional probabilities between variables. Currently, BN has been used in modeling causal relationships in many fields and has demonstrated advantages in causal interpretation, including in the fields such as hydrology and ecology (Chan et al., 2010; Keshtkar et al., 2013; Milns et al., 2010; Pollino et al., 2007; Shi et al., 2021a, b; Trifonova et al., 2015). However, BN has rarely been used in the study of the attribution of changes in ecosystem function. Therefore, this study used BN to attribute the controls of climate and plant traits to ecosystem function by quantifying the causal relationships involved. The data used are from a previous study (Migliavacca et al., 2021) which extracted ecosystem function, climate, and plant trait variables for FLUXNET flux stations. The construction of the causal structure of BN referred to the previous expert knowledge of this system (Reichstein et al., 2014). Further, by comparing BN-based attribution analysis, linear correlation analysis, and RF-based IMP reported by the previous study (Migliavacca et al., 2021), we investigated the adding-values of using BN for causal analysis and discussed its prospects in this paper.

2 Methodology

2.1 Data

The used variables (Table 1) include the carbon and water fluxes of the FLUXNET flux tower sites and the ecosystem function variables derived from them, and information on the corresponding climate variables as well as plant traits:

a) Ecosystem function variables: underlying Water Use Efficiency (uWUE), maximum evapotranspiration (ETmax), maximum surface conductance (GSmax), maximum net CO_2 uptake of the ecosystem (NEPmax), Gross Primary Productivity at light saturation (GPPsat), Mean basal ecosystem respiration at a reference temperature of 15 °C (Rb~~).~~), and apparent carbon-use efficiency (aCUE).

b) Plant trait variables: ecosystem scale foliar nitrogen concentration (Nmass), Maximum Leaf Area Index (LAImax), Maximum vegetation height (Hc~~), Aboveground Biomass (AGB).~~). Of the total 202 sites (Migliavacca and Musavi, 2021), 101 sites have Nmass data, 153 sites have LAImax data, and 199 sites have Hc data. Only 98 have data on all these three plant trait variables.

c) ~~Mean~~Climate variables: mean incoming shortwave radiation (SWin), Mean temperature (Tair), Mean Vapor Pressure Deficit (VPD), Mean annual precipitation (P), and cumulative soil water index (CSWI).

These data have different producing processes, including those calculated from flux data, site records, extracted from remote sensing data, etc. The ~~specific~~detailed calculation methods can be found in the ~~previous study~~ref. (Migliavacca et al., 2021).

Table 1. The variables used and the discretization of their values in BN.

Variable node	Definition and units	Type	~~Approach (Migliavacca et al., 2021)~~Approach (Migliavacca et al., 2021)	Discretization in BN (equal quantile thresholds~~ for classifications:~~ 0%, 33.33%, 66.67%, and 100% percentile values)
uWUE	underlying Water Use Efficiency [gC kPa^0.5 kgH_2O^{-1}]	Ecosystem function	~~calculated from GPP, VPD, ET~~It was calculated from GPP, VPD, and ET (Zhou et al., 2014). The median of the half-hourly retained uWUE values was used for each site. It was further filtered by the following conditions: (i) SWin > 200 W m^{-2}; (ii) no precipitation event for the last 24 hours, when precipitation data are available; and (iii) during the growing season: daily GPP > 30% of its seasonal amplitude.	0.~~0~~068, 2.~~5~~51, 3.18, 5~~, 5.5~~.332
ETmax	maximum evapotranspiration in the growing season [mm]	Ecosystem function	ETmax was computed as the 95th percentile of ET in the growing season. It was also filtered by the same filtering applied to the uWUE calculation.	0.~~05~~059, 0.~~15~~17, 0.~~30~~23, 0.~~45~~423
GSmax	maximum surface conductance [m s^{-1}]	Ecosystem function	GSmax was computed by inverting the Penman-Monteith equation after calculating the aerodynamic conductance. The 90th percentile of the half-hourly GS of each site was calculated and used as the GSmax of each site.	0.0013, 0.~~01~~0077, 0.~~02~~0123, 0.~~06~~0566
NEPmax	maximum net CO2 uptake of the ecosystem [umol CO_2 m^{-2} s^{-1}]	Ecosystem function	NEPmax was computed as the 90th percentile of the half-hourly net ecosystem production in the growing season (when daily GPP is > 30% of the GPP amplitude).	~~0~~1.953, 15~~, 30, 45~~.3, 24.4, 42.82
GPPsat	Gross Primary Productivity at light saturation [umol CO_2 m^{-2} s^{-1}]	Ecosystem function	GPPsat was computed as the 90th percentile estimated from half-hourly data by fitting the hyperbolic light response curves~~.~~. The 90th percentile from the GPPsat estimates of each site was extracted.	~~0, 15, 30, 50~~3.042, 17.49, 27.74, 47.6
Rb	Mean basal ecosystem respiration at a reference temperature of 15 °C [umol CO_2 m^{-2} s^{-1}]	Ecosystem function	Rb was derived from night-time NEE measurements. For each site, the mean of the daily Rb value was computed.	0.144, 2~~, 4, ~~.07, 3.12, 10.67
~~Nmassa~~CUE	~~ecosystem scale foliar nitrogen concentration~~	~~Plant trait~~Ecosystem function	~~computed as the community weighted average of foliar N% of the major species at the site~~	-1.19, 0.~~5, 1.25, 2.~~4, 0~~, 4.5~~.74, 1

	~~[gN 100 g-1]~~apparent carbon-use efficiency		~~sampled at the peak of the growing season or gathered from the literature (Musavi et al., 2016, 2015; Fleischer et al., 2015; Flechard et al., 2020)~~aCUE was calculated by aCUE = 1- (Rb/GPP) and the median value of daily aCUE is used.	
Nmass~~LAImax~~	~~Maximum Leaf Area Index [m^2 m^{-2}]~~ecosystem scale foliar nitrogen concentration [gN 100 g^{-1}]	Plant trait	~~collected from the literature (Migliavacca et al., 2011; Flechard et al., 2020), the FLUXNET Biological Ancillary Data Management (BADM) product, and/or site principal investigators~~Nmass was computed as the community-weighted average of foliar N% of the major species at the site sampled at the peak of the growing season or gathered from the literature (Musavi et al., 2016, 2015; Fleischer et al., 2015; Flechard et al., 2020).	~~0, 3, 6, 13~~0.65, 1.15, 1.76, 4.44
LAImax~~Hc~~	Maximum ~~vegetation height [~~Leaf Area Index [m^2 m^{-2}]	Plant trait	~~collected from the literature (Migliavacca et al., 2011; Flechard et al., 2020), the BADM product, and/or site principal investigators~~LAImax was collected from the literature (Migliavacca et al., 2011; Flechard et al., 2020), the FLUXNET Biological Ancillary Data Management (BADM) product, and/or site principal investigators.	0~~,~~.17, 2.27, 4.5, ~~20, 80~~12.9
~~AGB~~Hc	~~Aboveground Biomass derived from the Globbiomass project [t DM ha^{-1}]~~Maximum vegetation height [m]	Plant trait	~~extracted from the satellite-based GlobBiomass dataset (Santoro et al., 2021)~~Hc was collected from the literature (Migliavacca et al., 2011; Flechard et al., 2020), the BADM product, and/or site principal investigators.	~~0, 50, 150, 350~~0.04, 1.7, 16.0, 80.1

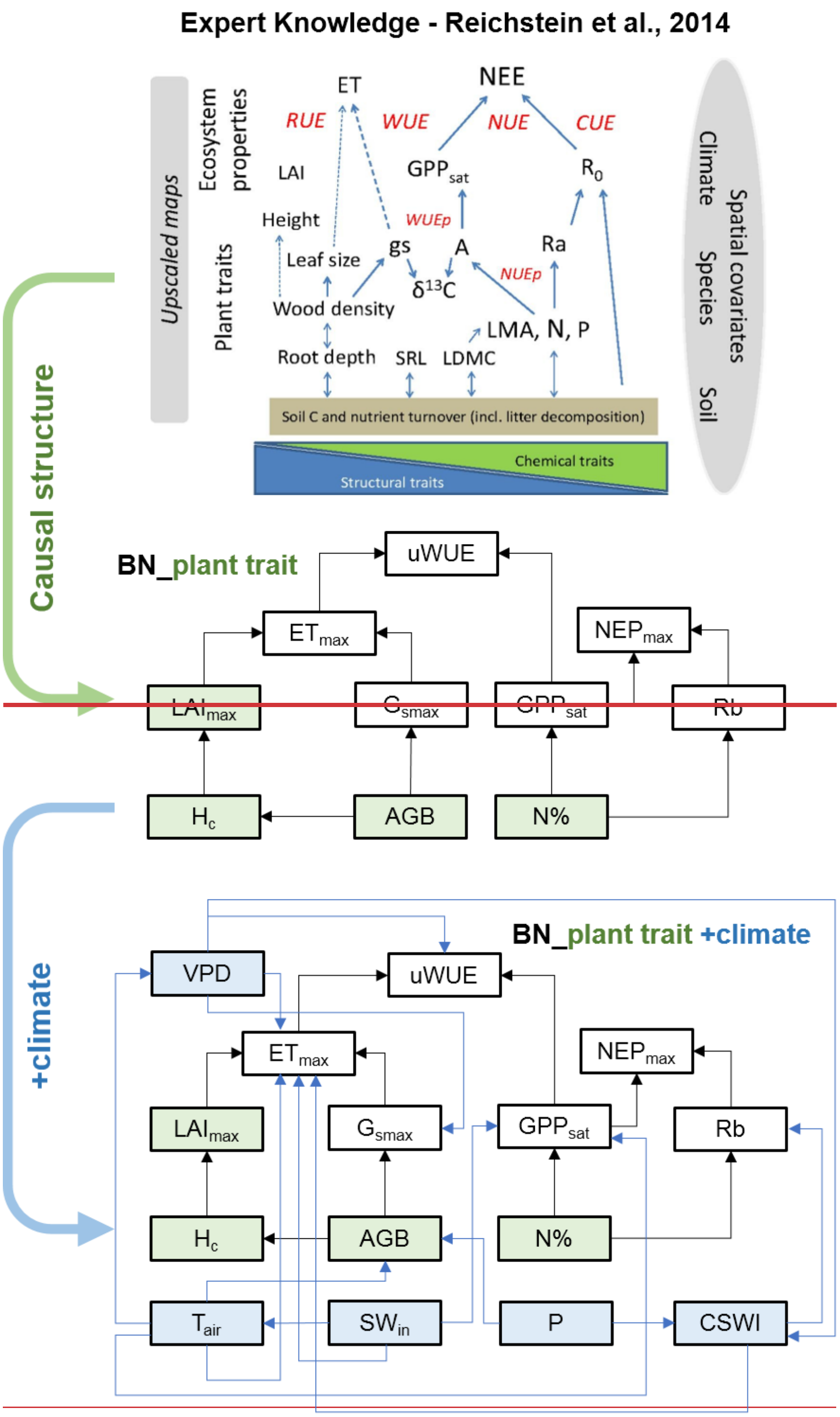
SWin	Mean incoming shortwave radiation [W m^{-2}]	Climate	SWin was from FLUXNET data.	~~50, 125, 200, 275~~54.43, 134.18, 182.44, 266.04
Tair	Mean temperature [degree C]	Climate	Tair was from FLUXNET data.	~~-12, 5, 15, 30~~-10.45, 6.62, 14.73, 28.1
VPD	Mean Vapor Pressure Deficit [hPa]	Climate	VPD was from FLUXNET data.	~~0, 4, 8, 27~~0.62, 3.38, 5.76, 26.08
P	Mean annual precipitation [cm/year]	Climate	P was from FLUXNET data.	~~0, 40, 80, 260~~5.51, 45.28, 79.29, 256.61
CSWI	cumulative soil water index	Climate-related soil water availability	~~computed as a measure of water availability (Nelson et al., 2018)~~CSWI was computed as a measure of water availability (Nelson et al., 2018).	~~-100, -20, 0, 5~~-93.49, -1.24, 2.01, 4.47

2.2 BN for analyzing causal relations

2.2.1 BN structures

Based on expert knowledge ~~(Reichstein et al., 2014)~~(Reichstein et al., 2014), we constructed the structure of BN containing the causal relationships between plant traits and ecosystem function variables: 'BN_plant_trait'. The causal links between the variables were referred to the relationship diagram in the upper part of Figure 1. Further, we added the climate variables and the corresponding causal relationships, expanding 'BN_plant_trait' to 'BN_plant_trait~~+~~_climate', which further incorporates the climate variables and their impacts on the system~~.~~ (Figure 1).

~~Each node is discretized for the BN compiling by the software Netica, while the selection of the thresholds for classifications (Table 1) is based on the distribution of the values of each node (Figure 2) and also the meanings of the thresholds. In this step, the three-level discretization (the distribution of a node is divided into three levels) for each node is applied given the limitation of the amount of training data.~~



Each node is discretized for the BN compiling by the software Netica. The equal quantile (Nojavan A. et al., 2017) three-level discretization (the distribution of nodes (Figure S1) is divided into three levels) for each node is applied by the discretization thresholds of 0%, 33.33%,66.67%, and 100% percentile values of the data distribution (Table 1) given the limitation of the amount of training data.

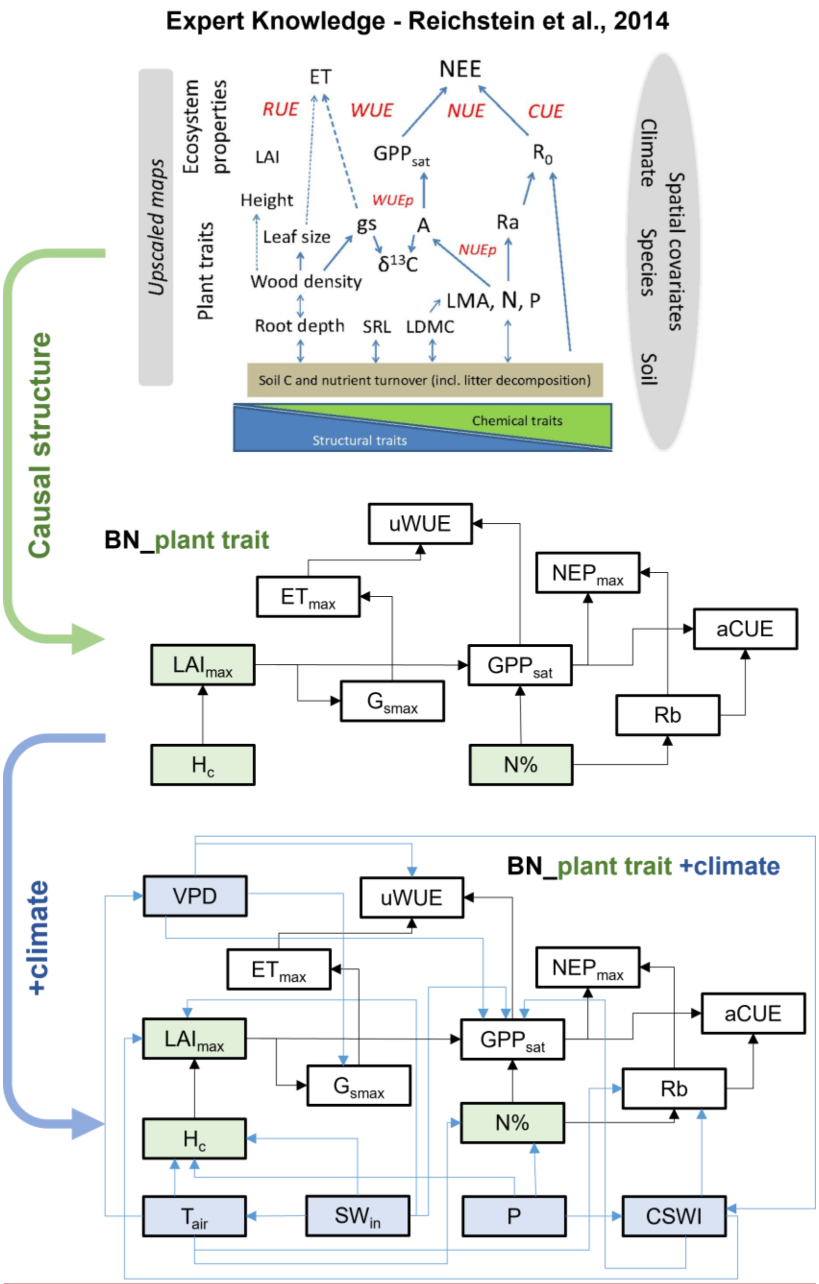
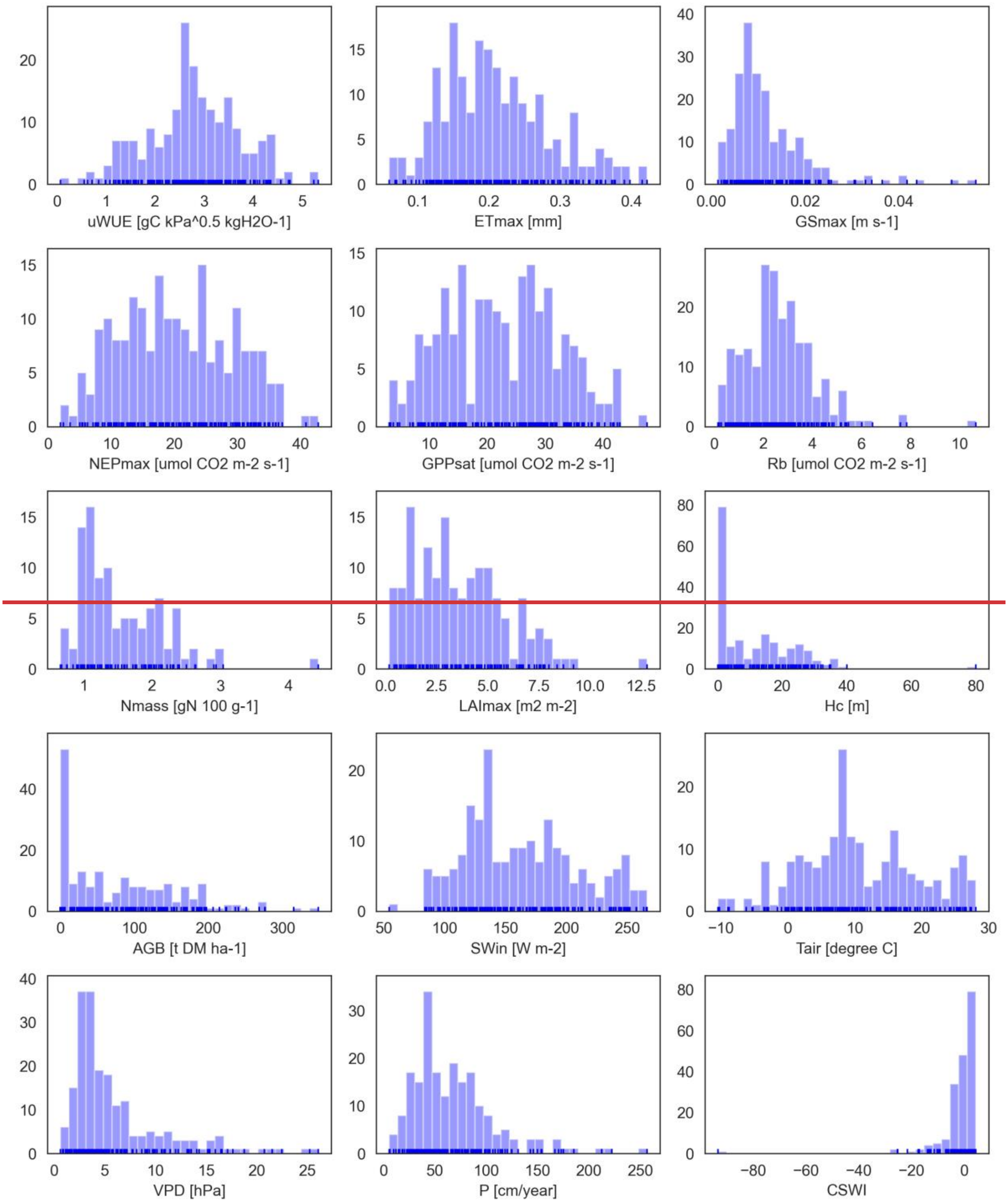


Figure 1. The structure of two Bayesian networks (BNs) for attribution of variations in ecosystem functions. ~~BN_pure in the lower left part assumes that all variables of plant traits (box in slight green) and climate (box in slight blue) directly affect the ecosystem function variables (box in white).~~ 'BN_plant_trait' in the median part incorporated the causal effects of plant traits (box in slight green) on ecosystem functions (box in white) from expert knowledge as the relation diagram on the upper part ~~(Reichstein et al., 2014).~~ ~~'BN_plant_trait+~~(Reichstein et al., 2014). 'BN_plant_trait_climate' in the lower part further incorporated the causal impacts of climate variables (box in light blue).



~~Figure~~ Table 2. ~~Distributions~~Explanation of ~~values of ecosystem functions and~~ the added causal links between climate ~~and~~variable nodes, plant trait ~~variables. The vertical axis indicates the number of flux stations~~nodes, and ecosystem function variable nodes in the BNs.

~~**2.2.2 Sensitivity analysis based on BN**~~

Casual links	Explanation	References

Parent node	Child node		
VPD	uWUE	$uWUE = GPP \cdot VPD^{0.5}/ET$	(Zhou et al., 2014)
VPD	GSmax	stomatal and surface conductance declines under an increase in VPD	(Grossiord et al., 2020; Wever et al., 2002)
VPD	GPPsat	leaf and canopy photosynthetic rates decline when atmospheric VPD increases due to stomatal closure	(Yuan et al., 2019; Konings et al., 2017)
Tair	VPD	higher air temperature corresponds to higher saturated water vapor pressure and can drive an increase in VPD	(Yuan et al., 2019)
Tair	Hc	the temperature limitation on canopy height variation	(Moles et al., 2009)
Tair	Nmass	increase in air temperature may decrease plant nitrogen concentration and leaf nitrogen content.	(Weih and Karlsson, 2001; Reich and Oleksyn, 2004)
Tair	Rb	temperature strongly influences Rb through the laws of thermodynamics	(Davidson and Janssens, 2006; Enquist et al., 2003; Brown et al., 2004)
SWin	LAImax	solar radiation affects vegetation conditions and phenology	(Günter et al., 2008; Liu et al., 2016; Borchert et al., 2015; Wagner et al., 2017)
SWin	Hc	solar radiation affects the distribution and composition of ecosystems through photosynthesis and the water cycle	(Borchert et al., 2015; Guisan and Zimmermann, 2000; Piedallu and Gégout, 2007)
SWin	GPPsat	solar radiation affects ecosystem productivity and plant growth	(Monteith, 1972; Borchert et al., 2015; Guisan and Zimmermann, 2000)
P	Hc	the hydraulic limitation hypothesis on canopy height variation	(Moles et al., 2009; Ryan and Yoder, 1997; Koch et al., 2004)
P	Nmass	leaf nitrogen concentration per unit mass may decrease with increasing precipitation	(Santiago and Mulkey, 2005; Wright and Westoby, 2002)
CSWI	LAImax	soil moisture affects vegetation conditions	(Patanè, 2011)
CSWI	Rb	soil moisture affects the temperature dependence of ecosystem respiration	(Xu et al., 2004; Flanagan and Johnson, 2005; Wen et al., 2006)
CSWI	GPPsat	soil moisture can reduce GPP through ecosystem water stress	(Green et al., 2019)

2.2.2 BN evaluation and node sensitivity analysis

Based on the Bayesian network (BN), the joint impacts of multiple variables and their causal relations are analyzed. A BN can be represented by nodes X_1, X_2, X_3 to X_n and the joint distribution ~~(Pearl, 1985)~~(Pearl, 1985):

$$\mathrm{Pa}(X) = \mathrm{Pa}(X_1, X_2, \dots, X_n) = \prod_{i=1}^{n} \mathrm{Pa}\big(X_i \big| \mathrm{pa}(X_i)\big) \tag{1}$$

where $\mathrm{pa}(X_i)$ is the probability of the parent node X_i. Expectation-maximization (Moon, 1996) is used to address the data with missing values and then compile the BN.

We used k-fold cross-validation to verify the reliability of the BN. The k-fold approach has been widely used in previous studies for the validation of BNs (Marcot, 2012). In this study, k is set as 10 as commonly used (Marcot and Hanea, 2021). We choose ETmax, GPPsat, and NEPmax for cross-validation of accuracy, and the predicted status (status with the highest probability bar value) of the nodes will be compared with the actual status and the classification accuracy will be calculated.

Sensitivity analysis is used for the evaluation of the strength of the causal relations between nodes based on mutual information (MI). MI is calculated as the entropy reduction of the child node resulting from changes found at the parent node (Shi et al., 2020a):

$$\mathrm{MI} = H(Q) - H(Q|F) = \sum_q \sum_f P(q, f) \log_2 \left(\frac{P(q,f)}{P(q)P(f)}\right) \tag{2}$$

where H represents the entropy, Q represents the target node, F represents the set of other nodes and q and f represent the status of Q and F. In this study, we assessed the sensitivity of ecosystem function variables to climate and plant trait variables.

2.2.3 Comparing different approaches used for attribution analysis

Further, to clarify the adding-values of considering causality in the attribution analysis of controls on ecosystem functions, the results of the BN-based sensitivity analysis (BN_sens) were compared with the other two approaches. They are the results of the absolute values of additional linear correlation analysis (linear_corr) in this study and the findings from the ref. (Migliavacca et al., 2021) using RF feature importance (RF_imp). BN_sens and linear_corr directly measure the effects of plant traits and climate variables on ecosystem function variables, while RF_imp measures their effects on the three principal components (PC1, PC2, and PC3) of ecosystem function variables, which were reported as the three major axes of ecosystem function by the ref. (Migliavacca et al., 2021). It was obtained from principal component analysis of 12 ecosystem function variables which included the six variables uWUE, ETmax, GSmax, NEPmax, GPPsat, and Rb used in the methods BN_sens and linear_corr. The first axis (PC1) explains 39.3% of the variance and is dominated by maximum ecosystem productivity properties, as indicated by the loadings of GPPsat and NEPmax, and maximum evapotranspiration (ETmax). The second axis (PC2) explains 21.4% of the variance and refers to water-use strategies as shown by the loadings of water-use efficiency metrics, evaporative fraction, and GSmax. The third axis (PC3) explains 11.1% of the variance and includes key attributes that reflect the carbon-use efficiency of ecosystems. PC3 is dominated by apparent carbon-use efficiency, basal ecosystem respiration (Rb), and the amplitude of evaporative fraction (Migliavacca et al., 2021).

~~Sensitivity analysis is used for the evaluation of the strength of the causal relations between nodes based on mutual information (MI). MI is calculated as the entropy reduction of the child node resulting from changes found at the parent node (Shi et al., 2020a):~~

$$\mathrm{MI} = \mathrm{H(Q)\text{-}H(Q|F)} = \sum_{q} \sum_{f} \mathrm{P(q,\ f)} \log_2 \left(\frac{\mathrm{P(q,f)}}{\mathrm{P(q)P(f)}}\right) \quad (2)$$

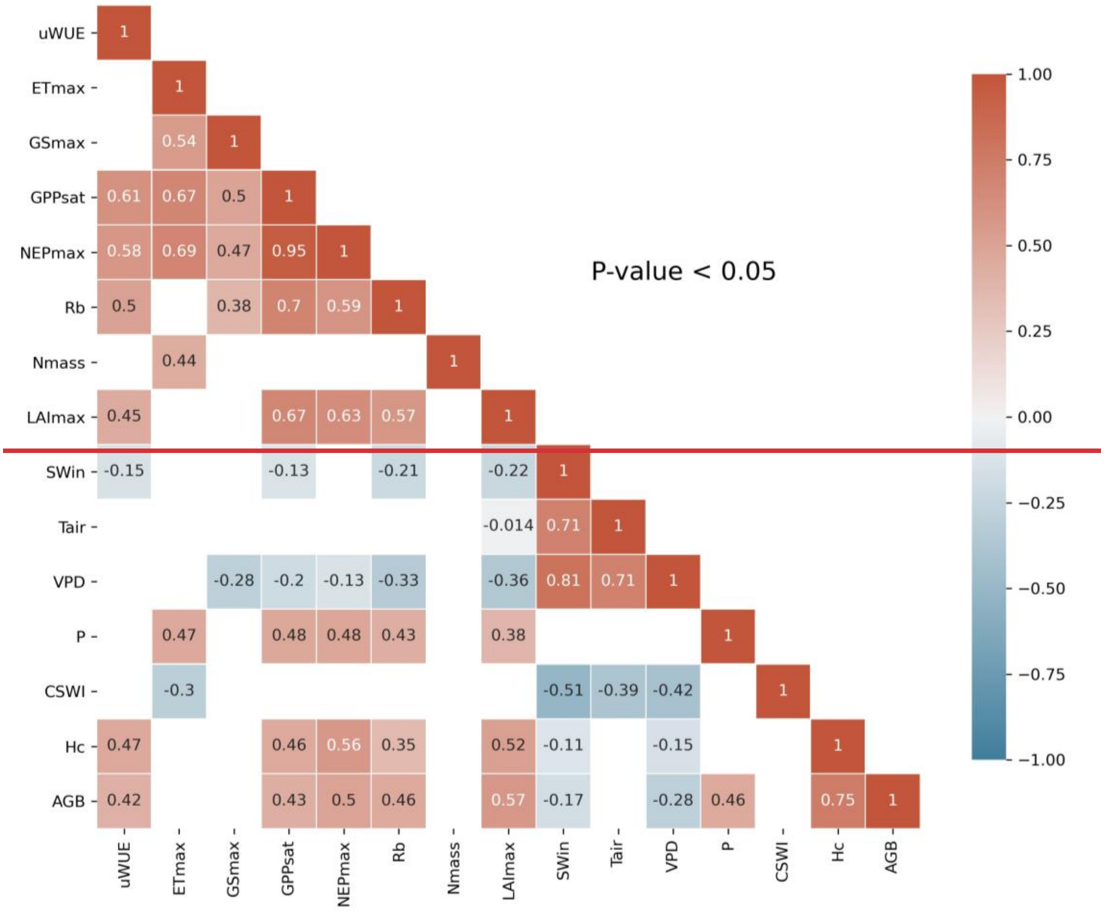
~~where H represents the entropy, Q represents the target node, F represents the set of other nodes and q and f represent the status of Q and F.~~

~~In this study, we assessed the sensitivity of ecosystem function variables to climate and plant trait variables. Further, to clarify the adding-values of considering causality in the attribution analysis of controls on ecosystem functions, the results of the BN-based sensitivity analysis were compared with the results of additional linear correlation analysis and the previous study using RF (Migliavacca et al., 2021) without considering the causality by comparing the ranking of MI, IMP (the feature importance metric of RF), and Pearson correlation coefficients (the metric of linear correlation analysis) of climate and plant trait variables and their differences in the results of the three methods. Although six ecosystem function variables were directly used in this study, the target variables of the RF-based approach were the first three principal components (PC): PC1, PC2, and PC3 (Migliavacca et al., 2021) of the 12 ecosystem variables (including the six variables selected in this study), the connotations of the target variables were relatively consistent between them.~~

3 Results

3.1 Correlation analysis

Linear correlation analysis of the variables (Figure ~~3~~2) showed significant ($P < 0.05$) linear correlations between the ecosystem function variables and some of the climate and plant trait variables. SWin~~,~~ and VPD~~, and~~ showed negative correlations with these ecosystem function variables. LAImax/ Hc showed significant positive relationships with most of the ecosystem function variables and significant negative relationships with SWin and VPD. Nmass only showed a positive relationship with ETmax. In addition, the majority of the ecosystem function variables showed significant ($P < 0.05$) positive correlations with each other.


P-value < 0.05
uWUE
ETmax
GSmax
GPPsat
NEPmax
Rb
Nmass
LAImax
SWin
Tair
VPD
P
CSWI
Hc
AGB

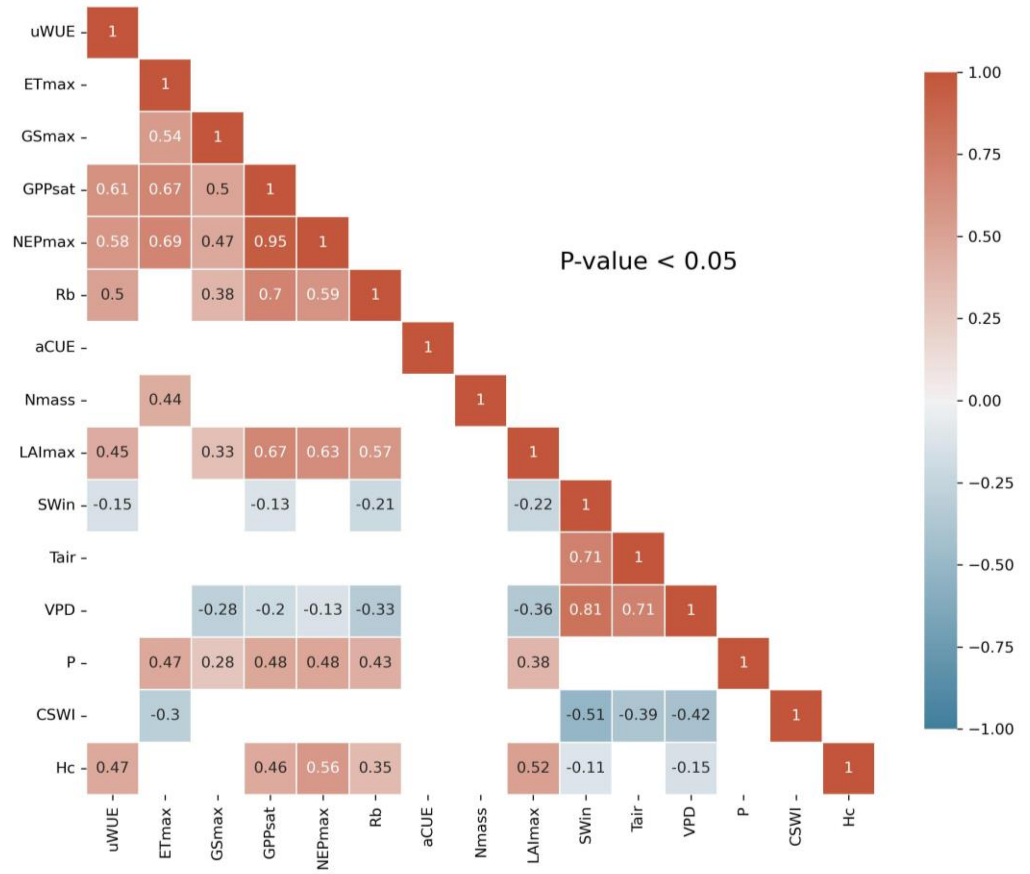
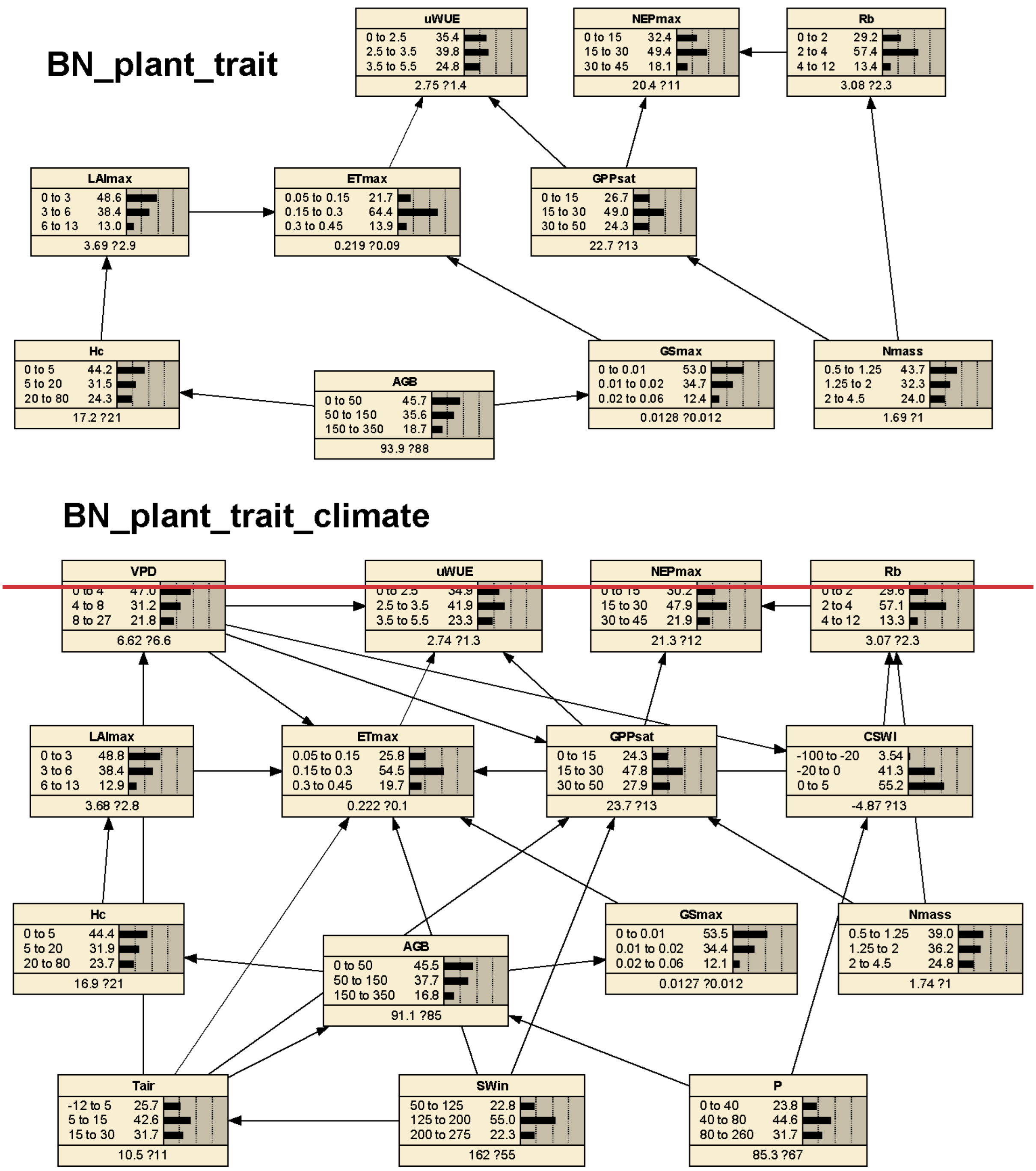

P-value < 0.05
uWUE
ETmax
GSmax
GPPsat
NEPmax
Rb
aCUE
Nmass
LAImax
SWin
Tair
VPD
P
CSWI
Hc

Figure ~~3~~2. Correlation coefficient matrix of ecosystem~~ service~~ functions and climate and plant trait variables for FLUXNET sites. Only correlation coefficients with p-values less than 0.05 level of significance ~~are~~is shown.

3.2 BN-based analysis

We compiled two different BNs (i.e., BN_plant_trait and BN_plant_trait_climate) (Figure ~~4~~3) and found that the probability distributions of the values of the common nodes (ecosystem function and plant trait variable nodes) differed ~~little between the two BNs. This indicates that the compilation was successful and that the inclusion of climate variables in BN_plant_trait_climate did not alter the fit of the local networks of ecosystem function and plant trait variables of BN_plant_trait.~~a little (e.g., in the probability distribution of LAImax, Hc, and Nmass) between the two BNs. Compared to BN_plant_trait, in BN_plant_trait_climate, the climate variables of sites with missing plant trait data forced the changes in the probability distributions of LAImax, Hc, and Nmass. In the EM algorithm, for sites with missing plant trait data, existing relationships (obtained from observations from other sites) between plant trait variables and climate variables are used in the data interpolation of plant trait variables. In BN_plant_trait_climate, the added linkages of climate variables to plant trait variables resulted in higher probability values of the low-value status of the plant trait variables.



The 10-fold cross-validation of the nodes ETmax, GPPsat, and NEPmax showed relatively high accuracy. The classification accuracy (Table S1) of the status of ETmax was 60.9%, the classification accuracy of the status of NEPmax was 84.2% and the classification accuracy of the status of GPPsat was 75.2%.

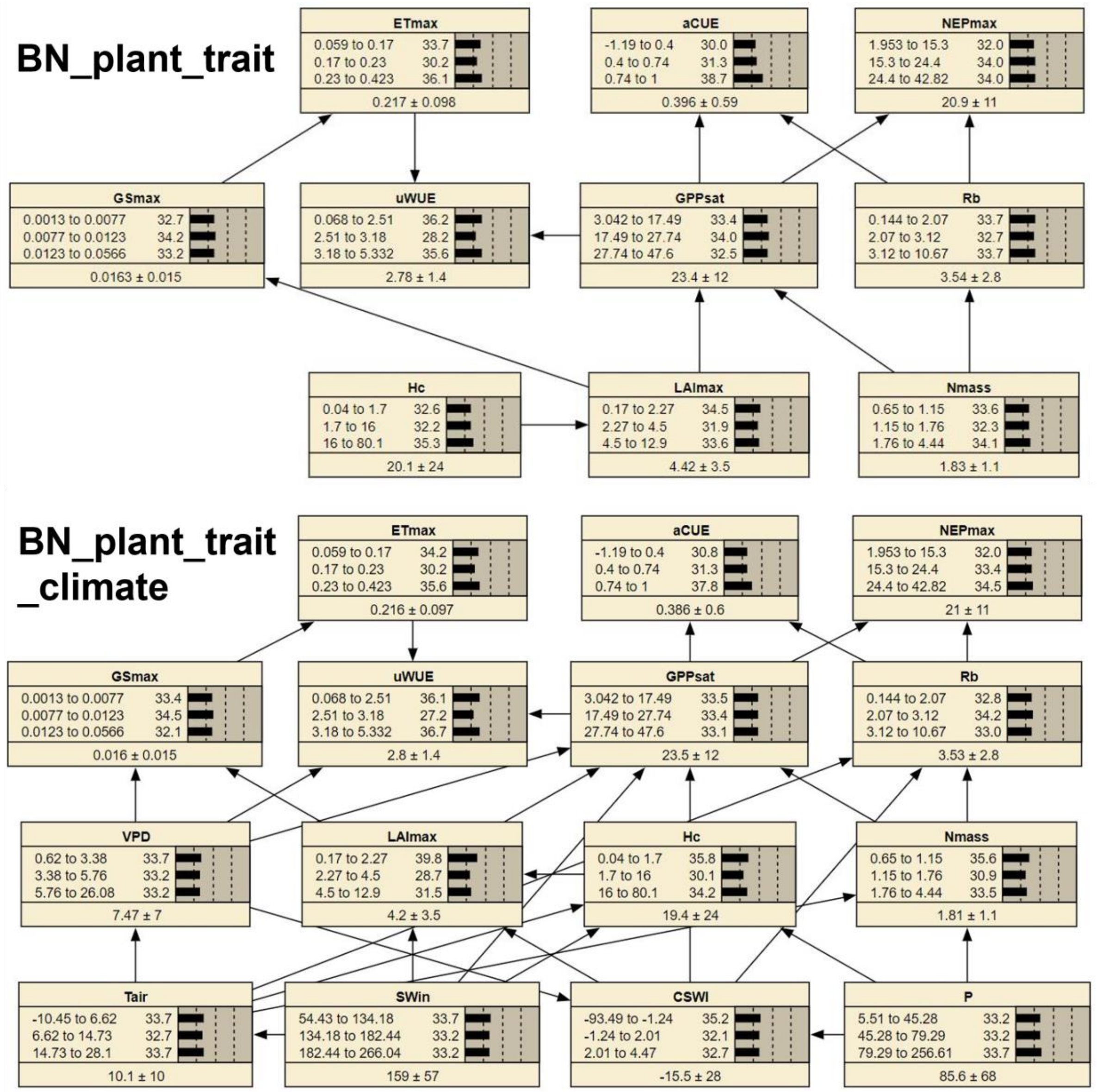


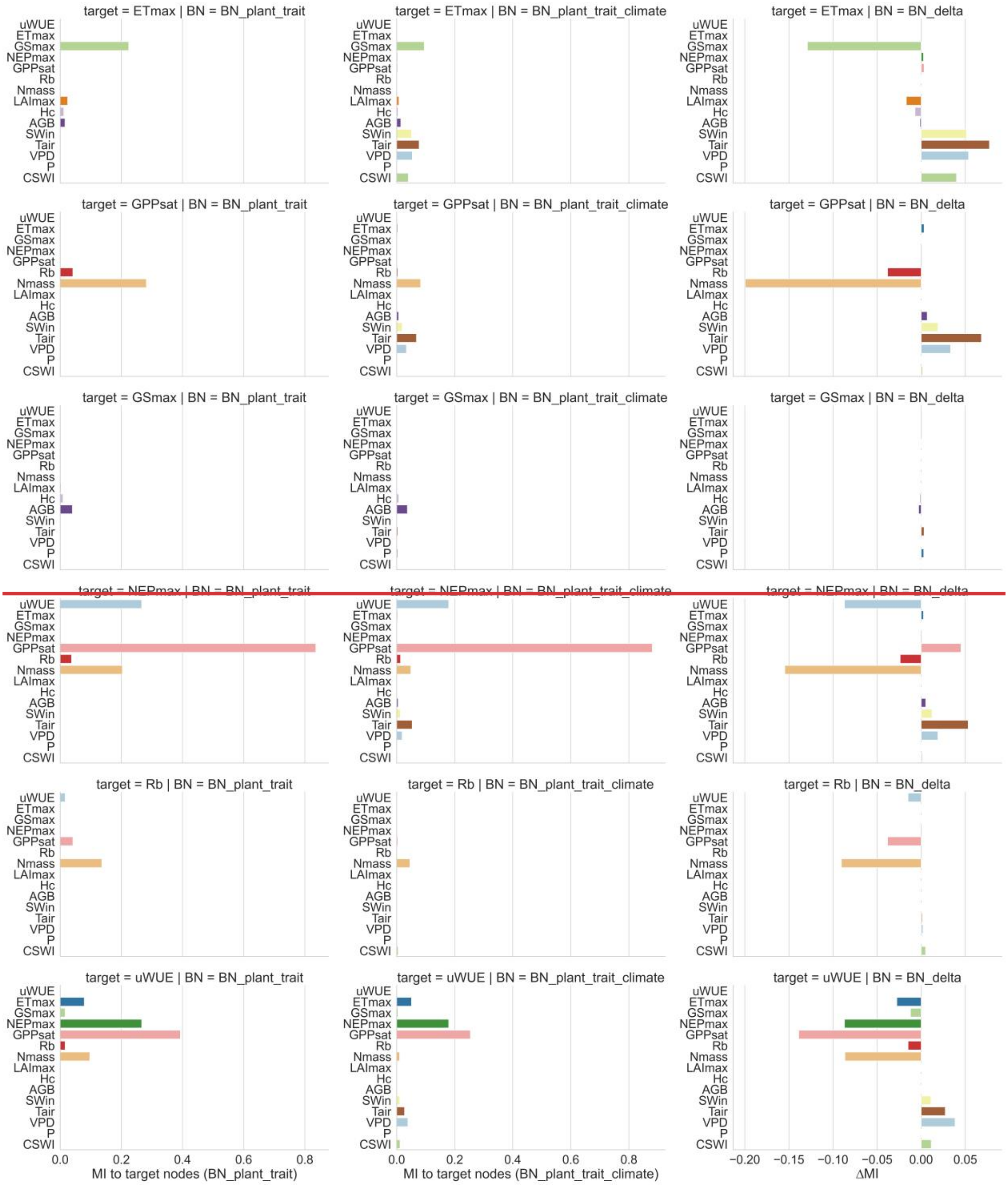
Figure ~~4.~~3. The compiled two BNs (‘BN_plant_trait’ and ‘BN_plant_trait_climate’). The bars of each node represent its probability distribution. At the bottom part of each node, the left and right side values of the '~~?~~±' are the mean and standard deviation of the distribution, respectively.

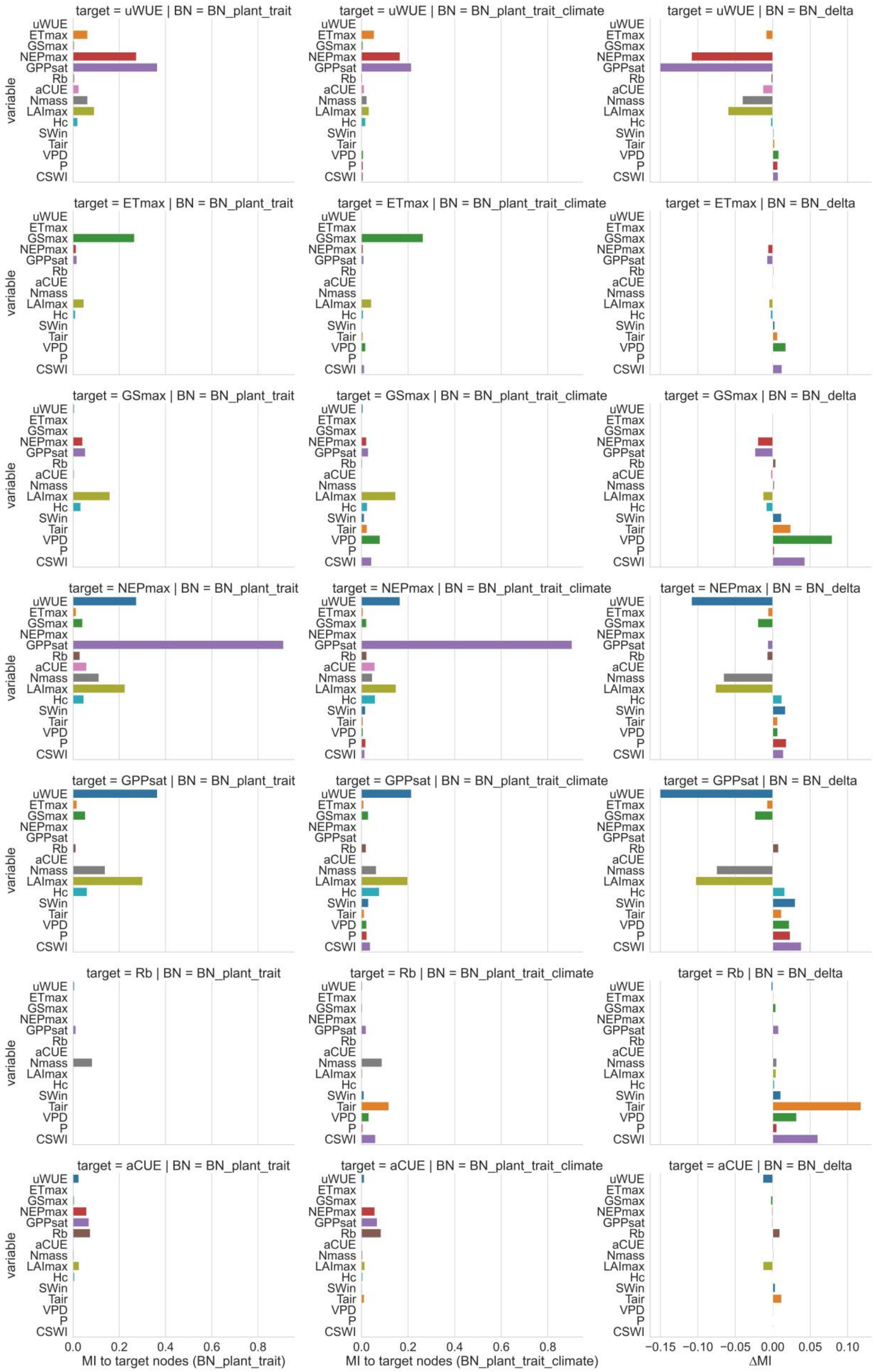
We performed sensitivity analyses (Figure ~~5~~4) on the ecosystem function variables in both BNs to assess their sensitivity to various climate and plant trait variables. We also calculated the difference in sensitivity MI between the two BNs (Figure ~~5~~4) to compare the change in sensitivity of ecosystem function to each variable after adding further climate variables to the plant trait variables only.

The sensitivity of different ecosystem function variables to plant traits and climate variables was highly variable in both BNs ~~(except for the similar pattern between GPPsat and NEPmax).~~. The magnitude of sensitivity of ecosystem ~~functional~~function nodes to plant traits and climate variables was related to whether these plant traits and climate variables were set as their parent nodes. In BN_plant_trait, for the carbon fluxes GPPsat and NEPmax, Nmass, and LAImax had ~~a~~ higher sensitivity due to Nmass and LAI being set as their parent

~~node~~nodes. For the water flux ETmax, it does not have high sensitivity to plant trait variables such as LAImax~~,~~ ~~Hc,~~ and ~~AGB~~Hc, although these plant trait variables are set as the parent nodes of ETmax. This indicates the difference in the strength of the control effects of plant traits on carbon and water fluxes.

In the sensitivity analysis of BN_plant_trait_climate, the sensitivity patterns of the ecosystem function variables changed as a result of the inclusion of climate variables and the change in causality they introduced. The sensitivity of the ecosystem function variables ~~(except GSmax and Rb)~~ to climate variables was significantly increased (especially for Tair~~ and~~, VPD~~). Their sensitivity to plant trait variables (e.g., Nmass and LAImax) decreased. Among the controls for ETmax and uWUE, climate variables showed a role beyond plant traits, while among the controls for GPPsat and NEPmax, the climate variable Tair also showed a significant role at a similar level to~~, and CSWI).~~Nmass.~~ The control of plant traits on ecosystem function in BN_plant_trait is also partially transformed into an indirect effect of climate variables by first controlling plant trait variables and then controlling ecosystem function. For example, in BN_plant_trait_climate, for ~~ETmax~~GPPsat, a decrease in the sensitivity of ~~ETmax~~GPPsat to LAImax and an increase in the sensitivity to Tair was observed after the ~~loop~~causal chain of Tair ~~controlling~~influencing Hc, LAImax, and then ~~ETmax~~GPPsat was set. This can be explained by the fact that higher temperatures promote vegetation growth and thus may increase LAImax, which then indirectly ~~contributes to~~alters the ~~increase in ETmax.~~probability distribution of the GPPsat node. In previous studies based on statistical methods that did not consider the ~~systematic~~chain causality, this indirect control on ~~ETmax~~GPPsat from Tair may have been included in the contribution of LAImax to ~~ETmax~~GPPsat. Similarly, a chain causality of P by first affecting Nmass and then indirectly GPPsat was also found. However, the effect of P by first affecting Hc, LAImax, and then indirectly affecting ETmax and GSmax appears to be not large.

333 
target = ETmax | BN = BN_plant_trait
target = ETmax | BN = BN_plant_trait_climate
target = ETmax | BN = BN_delta
target = GPPsat | BN = BN_plant_trait
target = GPPsat | BN = BN_plant_trait_climate
target = GPPsat | BN = BN_delta
target = GSmax | BN = BN_plant_trait
target = GSmax | BN = BN_plant_trait_climate
target = GSmax | BN = BN_delta
target = NEPmax | BN = BN_plant_trait
target = NEPmax | BN = BN_plant_trait_climate
target = NEPmax | BN = BN_delta
target = Rb | BN = BN_plant_trait
target = Rb | BN = BN_plant_trait_climate
target = Rb | BN = BN_delta
target = uWUE | BN = BN_plant_trait
target = uWUE | BN = BN_plant_trait_climate
target = uWUE | BN = BN_delta
uWUE
ETmax
GSmax
NEPmax
GPPsat
Rb
Nmass
LAImax
Hc
AGB
SWin
Tair
VPD
P
CSWI
0.0 0.2 0.4 0.6 0.8
MI to target nodes (BN_plant_trait)
MI to target nodes (BN_plant_trait_climate)
−0.20 −0.15 −0.10 −0.05 0.00 0.05
ΔMI



target = uWUE | BN = BN_plant_trait
target = uWUE | BN = BN_plant_trait_climate
target = uWUE | BN = BN_delta
target = ETmax | BN = BN_plant_trait
target = ETmax | BN = BN_plant_trait_climate
target = ETmax | BN = BN_delta
target = GSmax | BN = BN_plant_trait
target = GSmax | BN = BN_plant_trait_climate
target = GSmax | BN = BN_delta
target = NEPmax | BN = BN_plant_trait
target = NEPmax | BN = BN_plant_trait_climate
target = NEPmax | BN = BN_delta
target = GPPsat | BN = BN_plant_trait
target = GPPsat | BN = BN_plant_trait_climate
target = GPPsat | BN = BN_delta
target = Rb | BN = BN_plant_trait
target = Rb | BN = BN_plant_trait_climate
target = Rb | BN = BN_delta
target = aCUE | BN = BN_plant_trait
target = aCUE | BN = BN_plant_trait_climate
target = aCUE | BN = BN_delta
variable
uWUE
ETmax
GSmax
NEPmax
GPPsat
Rb
aCUE
Nmass
LAImax
Hc
SWin
Tair
VPD
P
CSWI
0.0 0.2 0.4 0.6 0.8
MI to target nodes (BN_plant_trait)
0.0 0.2 0.4 0.6 0.8
MI to target nodes (BN_plant_trait_climate)
−0.15 −0.10 −0.05 0.00 0.05 0.10
ΔMI

Figure ~~5~~4. Sensitivity of ecosystem function variables to other variables in different networks based on mutual information (MI). The left column is the sensitivity analysis of BN_plant_trait, the middle column is the sensitivity analysis of BN_plant_trait_climate, and the right column is the difference between the reported sensitivity of BN_plant_trait_climate and the sensitivity of BN_plant_trait. For BN_plant_trait, the MI values of climate variables to ecosystem function variables are all 0 because they do not contain climate variables. For each ecosystem function in these two BNs, its sensitivity to its child node is not shown (set as 0) because child nodes are not considered causal variables and thus are not evaluated in the attribution.

3.3 Comparing results from RF-based, BN-based analysis, and correlation analysis

All three methods show the importance of the plant trait variables in explaining the variation of various ecosystem function variables (Figure 5). LAImax was the most important of the three methods in explaining the variation of maximum ecosystem productivity properties (corresponding to PC1). In contrast to the results of the other two methods, in linear_corr, SWin and VPD were the least important, while P was more important. Comparing RF_imp and BN_sens, the overall pattern of importance is similar, but there are differences. For water-use strategies (corresponding to PC2), Hc is ranked first and LAI last in RF_imp, but in BN_sens, LAI is slightly more important than Hc. In linear_corr, Hc and LAI are of similar importance. For PC3, VPD ranks first and is more important than Tair in RF_imp. But in BN_sens, Tair is more important than VPD. Among the three moisture-related climate variables (i.e., VPD, P, and CSWI), CSWI appears to be the least important in RF_imp but is comparable to VPD in BN_sens.

Given the limitations of RF_imp in responding to the correlated variables (Strobl et al., 2008), the difference between the significance of VPD and CSWI reported by RF_imp may be overestimated. For the ecosystem functions related to water-use strategies, the difference between LAImax and Hc reported by BN_sens is also much smaller than the difference reported by RF_imp. It implied that, with the causality relation between correlated variables constructed, BN_sens reduced the uncertainty in quantifying the importance of correlated variables.

	Methods	Nmass	LAImax	Hc	SWin	Tair	VPD	P	CSWI
PC1	RF_imp	10.80%	16.60%	14.50%	7.60%	9.10%	11.70%	6.70%	4.00%
PC2	RF_imp	5.10%	4.50%	14.90%	10.70%	11.20%	7.40%	9.00%	8.30%
PC3	RF_imp	7.00%	2.80%	5.40%	9.30%	8.00%	15.40%	6.50%	4.90%
GPPsat	BN_sens	0.0635	0.1980	0.0766	0.0299	0.0116	0.0221	0.0232	0.0380
NEPmax	BN_sens	0.0464	0.1482	0.0588	0.0168	0.0064	0.0065	0.0181	0.0142
ETmax	BN_sens	0.0006	0.0424	0.0076	0.0028	0.0063	0.0174	0.0006	0.0122
uWUE	BN_sens	0.0228	0.0321	0.0174	0.0012	0.0023	0.0080	0.0066	0.0072
GSmax	BN_sens	0.0022	0.1464	0.0246	0.0115	0.0239	0.0793	0.0019	0.0429
Rb	BN_sens	0.0880	0.0043	0.0021	0.0106	0.1177	0.0317	0.0053	0.0602
aCUE	BN_sens	0.0049	0.0138	0.0056	0.0033	0.0117	0.0009	0.0004	0.0007
GPPsat	linear_corr		0.67	0.46	0.13		0.20	0.48	
NEPmax	linear_corr		0.63	0.56			0.13	0.48	
ETmax	linear_corr	0.44						0.47	0.30
uWUE	linear_corr		0.45	0.47	0.15				
GSmax	linear_corr						0.28		
Rb	linear_corr		0.57	0.35	0.21		0.33	0.43	
aCUE	linear_corr								

Figure 5. Comparisons of relationships of ecosystem functional variables to plant traits and climate variables in different analyses. Method RF_imp is Random forest variable importance (Migliavacca et al., 2021) (see Methodology section). Method linear_corr is Linear correlation analysis with the absolute values of Pearson correlation coefficients (see Methodology section). Method BN_sens is a BN-based sensitivity analysis with sensitivity values MI reported. ~~To compare the differences between cause-based and non-cause-based attribution or contribution analyses, we compared the importance ranking of variables based on the RF-based IMP in the study of Migliavacca et al., the absolute values of the correlation coefficients from the correlation analysis in this study, and the values of MI from the BN-based sensitivity analysis.~~

~~Since plant traits such as LAImax, Hc, and AGB were not set as parent nodes of carbon fluxes such as NEPmax and GPPsat in BN in this study, the effects of~~ The values in each method group are in red for high values and in blue for low values.

~~LAImax, Hc, and AGB on carbon flux-related variables in ecosystem function were weaker in the BN-based sensitivity analysis than in the RF-based and correlation analyses. However, AGB did not show a significant linear correlation with GSmax in the correlation analysis, suggesting that its control effect on GSmax may be nonlinear but detected by both RF and BN-based attribution analyses. Of the meteorological variables, Tair~~

~~showed stronger control over ecosystem function variables in the BN-based attribution (compared to other climate and plant trait variables), implying that the RF-based imputation of IMP may have underestimated the role of Tair.~~

	Methods	Nmass	LAImax	Hc	AGB	SWin	Tair	VPD	P	CSWI
PC1	RF_imp	10.80%	16.60%	14.50%	15.50%	7.60%	9.10%	11.70%	6.70%	4.00%
PC2	RF_imp	5.10%	4.50%	14.90%	5.10%	10.70%	11.20%	7.40%	9.00%	8.30%
PC3	RF_imp	7.00%	2.80%	5.40%	10.70%	9.30%	8.00%	15.40%	6.50%	4.90%
uWUE	BN_sens	0.010	0.000	0.000	0.001	0.011	0.027	0.038	0.000	0.011
ETmax	BN_sens	0.000	0.008	0.005	0.014	0.051	0.077	0.054	0.001	0.040
GSmax	BN_sens	0.000	0.002	0.008	0.037	0.000	0.003	0.000	0.003	0.000
NEPmax	BN_sens	0.049	0.000	0.000	0.005	0.012	0.053	0.019	0.000	0.001
GPPsat	BN_sens	0.082	0.000	0.000	0.007	0.019	0.068	0.033	0.000	0.002
Rb	BN_sens	0.046	0.000	0.000	0.000	0.000	0.001	0.002	0.000	0.005
uWUE	linear_corr		0.45	0.47	0.42	0.15				
ETmax	linear_corr	0.44							0.47	0.30
GSmax	linear_corr							0.28		
NEPmax	linear_corr		0.63	0.56	0.50			0.13	0.48	
GPPsat	linear_corr		0.67	0.46	0.43	0.13		0.20	0.48	
Rb	linear_corr		0.57	0.35	0.46	0.21		0.33	0.43	

~~Figure 6. Comparisons of relationships of ecosystem functional variables to plant traits and climate variables in different analyses. Method RF_imp is Random forest variable importance (Migliavacca et al., 2021). Method linear_corr is Linear correlation analysis with the absolute values of Pearson correlation coefficients. Method BN_sens is a BN-based sensitivity analysis with sensitivity values MI reported. PC1, PC2, and PC3 are the first three major axes of ecosystem function reported in the study by Migliavacca et al. (Migliavacca et al., 2021) obtained from principal component analysis of 12 ecosystem function variables which including the six variables uWUE, ETmax, GSmax, NEPmax, GPPsat, and Rb used in this study (Method BN_sens and linear_corr in the lower part). The first axis (PC1) explains 39.3% of the variance and is dominated by maximum ecosystem productivity properties, as indicated by the loadings of GPPsat and NEPmax, and maximum evapotranspiration (ETmax). The second axis (PC2) explains 21.4% of the variance and refers to water-use strategies as shown by the loadings of water-use efficiency metrics, evaporative fraction, and GSmax. The third axis (PC3) explains 11.1% of the variance and includes key attributes that reflect the carbon-use efficiency of ecosystems. PC3 is dominated by apparent carbon-use efficiency, basal ecosystem respiration (Rb), and the amplitude of evaporative fraction (Migliavacca et al., 2021). The values in each row are in red for high values and in blue for low values, for rows with very few values, the color-based indication is not reliable in ranking the control effects of plant traits and climate variables.~~

4 Discussions

~~Previous studies of 'climate-plant trait-ecosystem function' relationships have predominantly used only non-causal statistical methods such as RF (Migliavacca et al., 2021). Based on BN, this study investigates the prospect of using causal networks to revisit and attribute the control of climate and plant trait changes to ecosystem function. Compared to traditional correlation analysis and machine learning methods, BN can~~

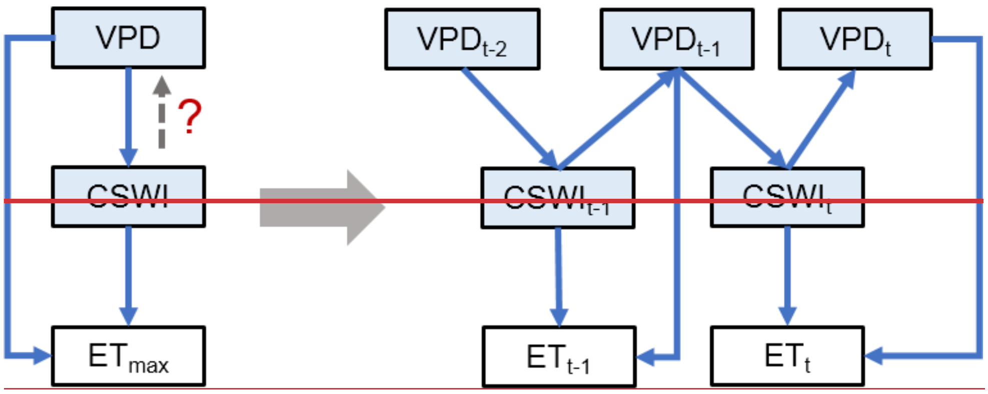
~~uncover the effects of causal relationships between variables. This causality discovery can improve on previous findings for studies of ecosystem-climate interactions.~~

Based on BN, this study investigates the prospect of using causal graphical models to revisit and attribute the control of climate and plant trait variations to ecosystem functions. Because of the inclusion of the constraints provided by expert knowledge ~~(Reichstein et al., 2014),~~(Reichstein et al., 2014) and other perceptions from many previous studies, BN-based attribution analysis is relatively reliable and can update our knowledge of the contribution of some teleconnection variables through causal chains. ~~However, since the structure of the expert knowledge graph did not connect LAImax, Hc, and AGB to GPP (Figure 1), LAImax, Hc, and AGB are not set as parent nodes of GPPsat in our BN, and thus the sensitivity of carbon fluxes to LAImax, Hc, AGB, etc. is 0 or close to 0. Therefore, in our BN, the causal controls of LAImax, Hc, and AGB on GPPsat are not shown although they are commonly thought to strongly influence GPPsat and NEPmax. This also demonstrates from another perspective the importance of a reasonable parent node in the attribution analysis using BN, where if a variable cannot be connected to the target variable through a causal loop, the sensitivity of the target variable to it may be low, and this will therefore affect the assessment of the strength of causality. If we want to explicitly measure the response of the target variable to the causality of a variable, it is indeed necessary to set up a causal link between them in BN.~~

~~In this study, it was found that the indirect impacts of some meteorological variables such as Tair may be underestimated in the attribution of ecosystem function when using a non-causal approach. This suggests the feasibility of quantifying indirect causal effects among various variables to help us gain a more systematic understanding.~~ The effective implementation of BN-based causal analysis may depend on the reliability of the causal relationships provided by expert knowledge (directional links between variables). We can establish the connection relationships and network structures between variables from expert knowledge and assign the specific quantification of the connection relationships (conditional probability tables) to ~~the~~ observations ~~and data (Shi et al., 2021a). In the future, we can revisit the linkages of ecosystem functions with climate and environmental systems using BN-based causal analysis to understand the strength and mechanisms of the relationships between direct, indirect, and remotely related effects of variables. Such a data-driven causal analysis framework provides more structured information about climate, plant traits, and ecosystem systems, thus making the data-driven approach more transparent and interpretable (compared to previous black-box models (Rudin, 2019)).~~(Shi et al., 2021a). If further combined with findings from process-based models, it is promising to significantly improve our understanding of the complex 'climate-plant trait-ecosystem function' relationships by comparing detailed relationships and structural influences between variables.

~~Besides, the BN in this study was mainly based on data averaged over multiple years, thus possibly partially underestimating the effect of temporal variations in the relationships between variables. Another limitation of the BN proposed above is that the causal relationships between variables are unidirectional, while it is difficult to represent interactions and feedback between variables (Marcot and Penman, 2019). In future studies, to address these two issues, BN based on temporal dynamics can be promising (Figure 7). By refining the interaction of temporal lags between variables, it is possible to incorporate not only temporal variation but also~~

~~control factors that attribute interactions and feedback between variables. For example, the interaction and feedback mechanisms of VPD, soil moisture, and ET with lag effects (Figure 7) and their impacts on ecosystems have attracted extensive interest from researchers (Anderegg et al., 2019; Humphrey et al., 2021; Lansu et al., 2020; Liu et al., 2020; Xu et al., 2022; Zhou et al., 2019), but conventional statistical methods have been ineffective in analyzing such relationships with both interactive causality and temporal lags. In contrast, the BN proposed here, which incorporates feedback effects and lagged effects that were common in climate-ecosystem relations (Lin et al., 2019), is potentially able to address this issue from a data-driven approach. When further combined with the findings of process-based models, our understanding of climate and ecosystem interactions and feedback and their mechanisms in time is hopefully deepened.~~



BN essentially factorizes the joint probability distribution among data variables into a series of conditional probability distributions (Ramazi et al., 2021), and the reliability of this approach relies on the setting of causal control relationships among nodes. Expert knowledge is thus critical in the construction of BNs, especially when modeling complex systems. In addition to the causal relationship between nodes, the meaning represented by each node, the data source/ approach, and the spatial and temporal resolution may also have impacts on the results. For example, in this study, for multiple water use efficiency-related variables in the ref. (Migliavacca et al., 2021), we chosed uWUE, and for Rb, we chosed the mean value of Rb. The results of BN-based analysis may vary if different representations or meanings of nodes are selected. The way the data of each variable is observed/ produced, the spatial and temporal resolution of the data, etc. can also affect the understanding of the role of these variables in the data-driven BN. Some variables may be very important in the attribution of actual ecosystem function variation, but their importance may be underestimated due to limitations in the inherent observational accuracy of their data, and differences in their spatial and temporal scales from other variables. In addition, some variables such as soil moisture may be difficult to obtain due to the lack of continuous site-scale long-term observations. Using the water balance method to calculate CSWI as a proxy may introduce errors. Since the CSWI calculation method relies on P, etc., the obtained relationship between P, CSWI, and other nodes may have contained empirical components. If the availability of measurements of some nodes is low, modelers should be cautious about the empirical dependencies with other nodes that may be included in the alternative data approaches. Thus, the alternative use of multiple derivatives of a variable and data generated by different methods for the construction of different BNs can help us to recognize how the uncertainty in the nodes and data can influence BN-based attribution findings. Different node discretization schemes may also affect the conditional probability table between nodes as well as the sensitivity (Nojavan A. et al., 2017). Other alternative discretization schemes with the commonly used three levels may also be effective, such as using 'mean-std' (mean minus 1 standard deviation) and 'mean+std' (mean plus 1 standard deviation) as discretization thresholds, which will result in a change in the relationship between BN nodes. And further if extreme values such as 5th and 95th pencentile are used in the node value discretization, it may be beneficial on quantifying the causal control of extreme conditions of nodes on other nodes.

When considering higher-order effects (Bairey et al., 2016), the relationships between plant traits, climate variables, and ecosystem function variables can be very complex. One variable may affect the relationship between two other variables rather than directly affecting these two variables (Bairey et al., 2016). BN may have limitations in directly analyzing such higher-order effects because BN requires the modeler to explicitly set direct causal relationships between nodes. To analyze the higher-order effects, we can add nodes that directly represent the relationship between the variables. For example, the correlation coefficient of two variables can be used as a node and this node is connected to other nodes in the BN so that the control effect of other nodes on this correlation coefficient can be explored. Such implements may be useful to deepen the impact of various higher order effects.

Besides, the BN in this study was mainly based on data averaged over multiple years, thus possibly partially underestimating the effect of temporal variations in the relationships between variables. Another limitation of the BN proposed above is that the causal relationships between variables are unidirectional, while it is difficult to represent interactions and feedback between variables (Marcot and Penman, 2019). In future studies, to address these two issues, BN based on temporal dynamics can be promising (Figure 6). By refining the interaction of temporal lags between variables, it is possible to incorporate not only temporal variation but also control factors that attribute interactions and feedback between variables. For example, the interaction and feedback mechanisms of VPD, soil moisture, and ET with lag effects (Figure 6) and their impacts on ecosystems have attracted extensive interest from researchers (Anderegg et al., 2019; Humphrey et al., 2021; Lansu et al., 2020; Liu et al., 2020; Xu et al., 2022; Zhou et al., 2019), but conventional statistical methods have been ineffective in analyzing such relationships with both interactive causality and temporal lags. In contrast, the BN proposed here, which incorporates feedback effects and lagged effects that were common in climate-ecosystem relations (Lin et al., 2019), is potentially able to address this issue from a data-driven approach. In the practical modeling, different periods of the same node may still be not independent. Therefore, the split scheme of such periods may be critical. For example, a period between two precipitation events can be treated as one sample, which can enhance independence between periods. Subsequently, a such period can be divided into smaller periods such as t, t-1, t-2, etc. to aggregate the node values to appropriate time scales. Thus one sample can represent the interaction relationship between variables with lags in this period. Finally, we can integrate records of such periods between two precipitation events from sites across different climate zones and biomes to build synthesis models for global analysis of such problems. If further combined with the findings of process-based models, our understanding of climate and ecosystem interactions and feedback and their mechanisms in time is hopefully deepened.

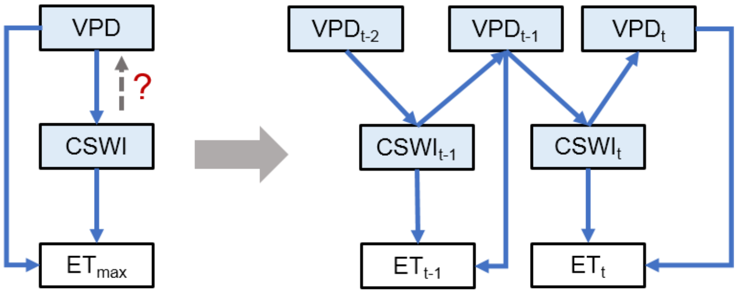


Figure ~~7~~6. The future BNs with the temporal causality further considered addressing the causality of the interaction between variables. The VPD-CSWI-ET relationship is used here as an example. t, t-1, and t-2 denote the current period, the last period, and the period before the last period, respectively. The network on the left only considers the effect of VPD on CSWI without considering the feedback of CSWI on the VPD. The network on the right characterizes the VPD-CSWI interaction with the feedback from CSWI at period t-1 to VPD at period t.

5 Conclusion

~~By emphasizing causality, based~~Based on BN, we revisited and attributed the contribution of climate and plant traits to global terrestrial ecosystem function. The major conclusions of this study include:

1. BN can be used for the quantification of causal relationships between complex ecosystems ~~and climatic and environmental systems~~in response to climate change and enables the analysis of indirect effects among variables.
2. ~~The control of ecosystem function~~Compared to BN, the feature importance difference between 'VPD and CSWI' and 'LAImax and Hc' reported by ~~climate~~ Random forests is higher and can be overestimated.
3. With the causality relation between correlated variables ~~(especially Tair and VPD) may have been underestimated in the past, and~~constructed, BN_sens can reduce the uncertainty in quantifying the importance of correlated variables.
4. ~~2.~~4. The understanding of the mechanism of indirect effects of climate variables on ecosystem function through plant traits ~~should be emphasized in future studies.~~can be deepened by the chain casuality quantification in BNs.
5. ~~3. Further inclusion of temporal information in BN holds promise for improving the analysis of lagged effects and interactions and feedback effects between variables.~~

Financial support

This research was supported by the National Natural Science Foundation of China (Grant No. U1803243), the Key projects of the Natural Science Foundation of Xinjiang Autonomous Region (Grant No. 2022D01D01), the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA20060302), and High-End Foreign Experts Project.

Author Contributions

HS and GL initiated this research and were responsible for the integrity of the work as a whole. HS performed formal analysis and calculations and drafted the manuscript. HS ~~were~~was responsible for the data collection and analysis. GL, PDM, TVdV, OH, and AK contributed resources and financial support.

Competing interests

The authors declare that they have no conflict of interest.

Code availability

The codes that were used for all analyses are available from the first author (shihaiyang16@mails.ucas.ac.cn) upon request.

Data availability

The data used in this study can be accessed by contacting the first author (shihaiyang16@mails.ucas.ac.cn) upon request.

References

Anderegg, W. R., Trugman, A. T., Bowling, D. R., Salvucci, G., and Tuttle, S. E.: Plant functional traits and climate influence drought intensification and land–atmosphere feedbacks, Proceedings of the National Academy of Sciences, 116, 14071–14076, 2019.

Bairey, E., Kelsic, E. D., and Kishony, R.: High-order species interactions shape ecosystem diversity, Nat Commun, 7, 1–7, https://doi.org/10.1038/ncomms12285, 2016.

Baldocchi, D.: Measuring fluxes of trace gases and energy between ecosystems and the atmosphere–the state and future of the eddy covariance method, Global change biology, 20, 3600–3609, 2014.

Barnes, M. L., Farella, M. M., Scott, R. L., Moore, D. J. P., Ponce-Campos, G. E., Biederman, J. A., MacBean, N., Litvak, M. E., and Breshears, D. D.: Improved dryland carbon flux predictions with explicit consideration of water-carbon coupling, Commun Earth Environ, 2, 1–9, https://doi.org/10.1038/s43247-021-00308-2, 2021.

Borchert, R., Calle, Z., Strahler, A. H., Baertschi, A., Magill, R. E., Broadhead, J. S., Kamau, J., Njoroge, J., and Muthuri, C.: Insolation and photoperiodic control of tree development near the equator, New Phytologist, 205, 7–13, 2015.

Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M., and West, G. B.: Toward a metabolic theory of ecology, Ecology, 85, 1771–1789, 2004.

Chan, T., Ross, H., Hoverman, S., and Powell, B.: Participatory development of a Bayesian network model for catchment-based water resource management, Water Resour. Res., 46, https://doi.org/10.1029/2009WR008848, 2010.

Chapin Iii, F. S., Zavaleta, E. S., Eviner, V. T., Naylor, R. L., Vitousek, P. M., Reynolds, H. L., Hooper, D. U., Lavorel, S., Sala, O. E., and Hobbie, S. E.: Consequences of changing biodiversity, Nature, 405, 234–242, 2000.

Davidson, E. A. and Janssens, I. A.: Temperature sensitivity of soil carbon decomposition and feedbacks to climate change, Nature, 440, 165–173, 2006.

Diaz, S. and Cabido, M.: Plant functional types and ecosystem function in relation to global change, Journal of Vegetation Science, 8, 463–474, https://doi.org/10.2307/3237198, 1997.

Enquist, B. J., Economo, E. P., Huxman, T. E., Allen, A. P., Ignace, D. D., and Gillooly, J. F.: Scaling metabolism from organisms to ecosystems, Nature, 423, 639–642, 2003.

Flanagan, L. B. and Johnson, B. G.: Interacting effects of temperature, soil moisture and plant biomass production on ecosystem respiration in a northern temperate grassland, Agricultural and Forest Meteorology, 130, 237–253, 2005.

Flechard, C. R., Ibrom, A., Skiba, U. M., de Vries, W., van Oijen, M., Cameron, D. R., Dise, N. B., Korhonen, J. F. J., Buchmann, N., Legout, A., Simpson, D., Sanz, M. J., Aubinet, M., Loustau, D., Montagnani, L., Neirynck, J., Janssens, I. A., Pihlatie, M., Kiese, R., Siemens, J., Francez, A.-J., Augustin, J., Varlagin, A., Olejnik, J., Juszczak, R., Aurela, M., Berveiller, D., Chojnicki, B. H., Dämmgen, U., Delpierre, N., Djuricic, V., Drewer, J., Dufrêne, E., Eugster, W., Fauvel, Y., Fowler, D., Frumau, A., Granier, A., Gross, P., Hamon, Y., Helfter, C., Hensen, A., Horváth, L., Kitzler, B., Kruijt, B., Kutsch, W. L., Lobo-do-Vale, R., Lohila, A., Longdoz, B., Marek, M. V., Matteucci, G., Mitosinkova, M., Moreaux, V., Neftel, A., Ourcival, J.-M., Pilegaard, K., Pita, G., Sanz, F., Schjoerring, J. K., Sebastià, M.-T., Tang, Y. S., Uggerud, H., Urbaniak, M., van Dijk, N., Vesala, T., Vidic, S., Vincke, C., Weidinger, T., Zechmeister-Boltenstern, S., Butterbach-Bahl, K., Nemitz, E., and Sutton, M. A.: Carbon–nitrogen interactions in European forests and semi-natural vegetation – Part 1: Fluxes and budgets of carbon, nitrogen and greenhouse gases from ecosystem monitoring and modelling, Biogeosciences, 17, 1583–1620, https://doi.org/10.5194/bg-17-1583-2020, 2020.

Fleischer, K., Wårlind, D., Van der Molen, M. K., Rebel, K. T., Arneth, A., Erisman, J. W., Wassen, M. J., Smith, B., Gough, C. M., and Margolis, H. A.: Low historical nitrogen deposition effect on carbon sequestration in the boreal zone, Journal of Geophysical Research: Biogeosciences, 120, 2542–2561, 2015.

Friedman, N., Geiger, D., and Goldszmidt, M.: Bayesian network classifiers, Machine learning, 29, 131–163, 1997.

~~Gregorutti, B., Michel, B., and Saint-Pierre, P.: Correlation and variable importance in random forests, Statistics and Computing, 27, 659–678, 2017.~~

Green, J. K., Seneviratne, S. I., Berg, A. M., Findell, K. L., Hagemann, S., Lawrence, D. M., and Gentine, P.: Large influence of soil moisture on long-term terrestrial carbon uptake, Nature, 565, 476–479, 2019.

Grimm, N. B., Chapin III, F. S., Bierwagen, B., Gonzalez, P., Groffman, P. M., Luo, Y., Melton, F., Nadelhoffer, K., Pairis, A., and Raymond, P. A.: The impacts of climate change on ecosystem structure and function, Frontiers in Ecology and the Environment, 11, 474–482, 2013.

de Groot, R. S., Wilson, M. A., and Boumans, R. M. J.: A typology for the classification, description and valuation of ecosystem functions, goods and services, Ecological Economics, 41, 393–408, https://doi.org/10.1016/S0921-8009(02)00089-7, 2002.

Grossiord, C., Buckley, T. N., Cernusak, L. A., Novick, K. A., Poulter, B., Siegwolf, R. T. W., Sperry, J. S., and McDowell, N. G.: Plant responses to rising vapor pressure deficit, New Phytologist, 226, 1550–1566, https://doi.org/10.1111/nph.16485, 2020.

Guisan, A. and Zimmermann, N. E.: Predictive habitat distribution models in ecology, Ecological modelling, 135, 147–186, 2000.

Günter, S., Stimm, B., Cabrera, M., Diaz, M. L., Lojan, M., Ordonez, E., Richter, M., and Weber, M.: Tree phenology in montane forests of southern Ecuador can be explained by precipitation, radiation and photoperiodic control, Journal of Tropical Ecology, 24, 247–258, 2008.

Humphrey, V., Berg, A., Ciais, P., Gentine, P., Jung, M., Reichstein, M., Seneviratne, S. I., and Frankenberg, C.: Soil moisture–atmosphere feedback dominates land carbon uptake variability, Nature, 592, 65–69, https://doi.org/10.1038/s41586-021-03325-5, 2021.

Jung, M., Reichstein, M., Ciais, P., Seneviratne, S. I., Sheffield, J., Goulden, M. L., Bonan, G., Cescatti, A., Chen, J., de Jeu, R., Dolman, A. J., Eugster, W., Gerten, D., Gianelle, D., Gobron, N., Heinke, J., Kimball, J., Law, B. E., Montagnani, L., Mu, Q., Mueller, B., Oleson, K., Papale, D., Richardson, A. D., Roupsard, O., Running, S., Tomelleri, E., Viovy, N., Weber, U., Williams, C., Wood, E., Zaehle, S., and Zhang, K.: Recent decline in the global land evapotranspiration trend due to limited moisture supply, Nature, 467, 951–954, https://doi.org/10.1038/nature09396, 2010.

Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., Anthoni, P., Besnard, S., Bodesheim, P., Carvalhais, N., Chevallier, F., Gans, F., S Goll, D., Haverd, V., Köhler, P., Ichii, K., K Jain, A., Liu, J., Lombardozzi, D., E M S Nabel, J., A Nelson, J., O'Sullivan, M., Pallandt, M., Papale, D., Peters, W., Pongratz, J., Rödenbeck, C., Sitch, S., Tramontana, G., Walker, A., Weber, U., and Reichstein, M.: Scaling carbon fluxes from eddy covariance sites to globe:

Synthesis and evaluation of the FLUXCOM approach, Biogeosciences, 17, 1343–1365, https://doi.org/10.5194/bg-17-1343-2020, 2020.

Keshtkar, A. R., Salajegheh, A., Sadoddin, A., and Allan, M. G.: Application of Bayesian networks for sustainability assessment in catchment modeling and management (Case study: The Hablehrood river catchment), Ecological Modelling, 268, 48–54, 2013.

Koch, G. W., Sillett, S. C., Jennings, G. M., and Davis, S. D.: The limits to tree height, Nature, 428, 851–854, 2004.

Konings, A., Williams, A., and Gentine, P.: Sensitivity of grassland productivity to aridity controlled by stomatal and xylem regulation, Nature Geoscience, 10, 284–288, 2017.

Lansu, E. M., van Heerwaarden, C., Stegehuis, A. I., and Teuling, A. J.: Atmospheric aridity and apparent soil moisture drought in European forest during heat waves, Geophysical Research Letters, 47, e2020GL087091, 2020.

Lin, C., Gentine, P., Frankenberg, C., Zhou, S., Kennedy, D., and Li, X.: Evaluation and mechanism exploration of the diurnal hysteresis of ecosystem fluxes, Agricultural and Forest Meteorology, 278, 107642, https://doi.org/10.1016/j.agrformet.2019.107642, 2019.

Liu, L., Gudmundsson, L., Hauser, M., Qin, D., Li, S., and Seneviratne, S. I.: Soil moisture dominates dryness stress on ecosystem production globally, Nature communications, 11, 1–9, 2020.

Liu, Q., Fu, Y. H., Zeng, Z., Huang, M., Li, X., and Piao, S.: Temperature, precipitation, and insolation effects on autumn vegetation phenology in temperate China, Global Change Biology, 22, 644–655, https://doi.org/10.1111/gcb.13081, 2016.

Madani, N., Kimball, J. S., Ballantyne, A. P., Affleck, D. L. R., van Bodegom, P. M., Reich, P. B., Kattge, J., Sala, A., Nazeri, M., Jones, M. O., Zhao, M., and Running, S. W.: Future global productivity will be affected by plant trait response to climate, Sci Rep, 8, 2870, https://doi.org/10.1038/s41598-018-21172-9, 2018.

Manning, P., Van Der Plas, F., Soliveres, S., Allan, E., Maestre, F. T., Mace, G., Whittingham, M. J., and Fischer, M.: Redefining ecosystem multifunctionality, Nature ecology & evolution, 2, 427–436, 2018.

Marcot, B. G.: Metrics for evaluating performance and uncertainty of Bayesian network models, Ecological modelling, 230, 50–62, 2012.

Marcot, B. G. and Hanea, A. M.: What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis?, Comput Stat, 36, 2009–2031, https://doi.org/10.1007/s00180-020-00999-9, 2021.

Marcot, B. G. and Penman, T. D.: Advances in Bayesian network modelling: Integration of modelling technologies, Environmental modelling & software, 111, 386–393, 2019.

Migliavacca, M. and Musavi, T.: Reproducible Workflow: The three major axes of terrestrial ecosystem function, https://doi.org/10.5281/zenodo.5153538, 2021.

Migliavacca, M., Reichstein, M., Richardson, A. D., Colombo, R., Sutton, M. A., Lasslop, G., Tomelleri, E., Wohlfahrt, G., Carvalhais, N., and Cescatti, A.: Semiempirical modeling of abiotic and biotic factors controlling ecosystem respiration across eddy covariance sites, Global Change Biology, 17, 390–409, 2011.

Migliavacca, M., Musavi, T., Mahecha, M. D., Nelson, J. A., Knauer, J., Baldocchi, D. D., Perez-Priego, O., Christiansen, R., Peters, J., Anderson, K., Bahn, M., Black, T. A., Blanken, P. D., Bonal, D., Buchmann, N., Caldararu, S., Carrara, A., Carvalhais, N., Cescatti, A., Chen, J., Cleverly, J., Cremonese, E., Desai, A. R., El-Madany, T. S., Farella, M. M., Fernández-Martínez, M., Filippa, G., Forkel, M., Galvagno, M., Gomarasca, U., Gough, C. M., Göckede, M., Ibrom, A., Ikawa, H., Janssens, I. A., Jung, M., Kattge, J., Keenan, T. F., Knohl, A., Kobayashi, H., Kraemer, G., Law, B. E., Liddell, M. J., Ma, X., Mammarella, I., Martini, D., Macfarlane, C., Matteucci, G., Montagnani, L., Pabon-Moreno, D. E., Panigada, C., Papale, D., Pendall, E., Penuelas, J., Phillips, R. P., Reich, P. B., Rossini, M., Rotenberg, E., Scott, R. L., Stahl, C., Weber, U., Wohlfahrt, G., Wolf, S., Wright, I. J., Yakir, D., Zaehle, S., and Reichstein, M.: The three major axes of terrestrial ecosystem function, Nature, 598, 468–472, https://doi.org/10.1038/s41586-021-03939-9, 2021.

Milns, I., Beale, C. M., and Smith, V. A.: Revealing ecological networks using Bayesian network inference algorithms, Ecology, 91, 1892–1899, https://doi.org/10.1890/09-0731.1, 2010.

Moles, A. T., Warton, D. I., Warman, L., Swenson, N. G., Laffan, S. W., Zanne, A. E., Pitman, A., Hemmings, F. A., and Leishman, M. R.: Global patterns in plant height, Journal of ecology, 97, 923–932, 2009.

Monteith, J. L.: Solar radiation and productivity in tropical ecosystems, Journal of applied ecology, 9, 747–766, 1972.

Moon, T. K.: The expectation-maximization algorithm, IEEE Signal processing magazine, 13, 47–60, 1996.

Musavi, T., Mahecha, M. D., Migliavacca, M., Reichstein, M., van de Weg, M. J., van Bodegom, P. M., Bahn, M., Wirth, C., Reich, P. B., and Schrodt, F.: The imprint of plants on ecosystem functioning: A data-driven approach, International Journal of Applied Earth Observation and Geoinformation, 43, 119–131, 2015.

Musavi, T., Migliavacca, M., van de Weg, M. J., Kattge, J., Wohlfahrt, G., van Bodegom, P. M., Reichstein, M., Bahn, M., Carrara, A., and Domingues, T. F.: Potential and limitations of inferring ecosystem photosynthetic capacity from leaf functional traits, Ecology and evolution, 6, 7352–7366, 2016.

Myers-Smith, I. H., Thomas, H. J. D., and Bjorkman, A. D.: Plant traits inform predictions of tundra responses to global change, New Phytologist, 221, 1742–1748, https://doi.org/10.1111/nph.15592, 2019.

Nelson, J. A., Carvalhais, N., Migliavacca, M., Reichstein, M., and Jung, M.: Water-stress-induced breakdown of carbon–water relations: indicators from diurnal FLUXNET patterns, Biogeosciences, 15, 2433–2447, 2018.

Nojavan A., F., Qian, S. S., and Stow, C. A.: Comparative analysis of discretization methods in Bayesian networks, Environmental Modelling & Software, 87, 64–71, https://doi.org/10.1016/j.envsoft.2016.10.007, 2017.

Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W., Poindexter, C., Chen, J., Elbashandy, A., Humphrey, M., Isaac, P., Polidori, D., Reichstein, M., Ribeca, A., van Ingen, C., Vuichard, N., Zhang, L., Amiro, B., Ammann, C., Arain, M. A., Ardö, J., Arkebauer, T., Arndt, S. K., Arriga, N., Aubinet, M., Aurela, M., Baldocchi, D., Barr, A., Beamesderfer, E., Marchesini, L. B., Bergeron, O., Beringer, J., Bernhofer, C., Berveiller, D., Billesbach, D., Black, T. A., Blanken, P. D., Bohrer, G., Boike, J., Bolstad, P. V., Bonal, D., Bonnefond, J.-M., Bowling, D. R., Bracho, R., Brodeur, J., Brümmer, C., Buchmann, N., Burban, B., Burns, S. P., Buysse, P., Cale, P., Cavagna, M., Cellier, P., Chen, S., Chini, I., Christensen, T. R., Cleverly, J., Collalti, A., Consalvo, C., Cook, B. D., Cook, D., Coursolle, C., Cremonese, E., Curtis, P. S., D'Andrea, E., da Rocha, H., Dai, X., Davis, K. J., Cinti, B. D., Grandcourt, A. de, Ligne, A. D., De Oliveira, R. C., Delpierre, N., Desai, A. R., Di Bella, C. M., Tommasi, P. di, Dolman, H., Domingo, F., Dong, G., Dore, S., Duce, P., Dufrêne, E., Dunn, A., Dušek, J., Eamus, D., Eichelmann, U., ElKhidir, H. A. M., Eugster, W., Ewenz, C. M., Ewers, B., Famulari, D., Fares, S., Feigenwinter, I., Feitz, A., Fensholt, R., Filippa, G., Fischer, M., Frank, J., Galvagno, M., et al.: The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data, Sci Data, 7, 225, https://doi.org/10.1038/s41597-020-0534-3, 2020.

Patanè, C.: Leaf Area Index, Leaf Transpiration and Stomatal Conductance as Affected by Soil Water Deficit and VPD in Processing Tomato in Semi Arid Mediterranean Climate, Journal of Agronomy and Crop Science, 197, 165–176, https://doi.org/10.1111/j.1439-037X.2010.00454.x, 2011.

Pearl, J.: Bayesian netwcrks: A model cf self-activated memory for evidential reasoning, in: Proceedings of the 7th Conference of the Cognitive Science Society, University of California, Irvine, CA, USA, 15–17, 1985.

Peaucelle, M., Bacour, C., Ciais, P., Vuichard, N., Kuppel, S., Peñuelas, J., Belelli Marchesini, L., Blanken, P. D., Buchmann, N., and Chen, J.: Covariations between plant functional traits emerge from constraining parameterization of a terrestrial biosphere model, Global ecology and biogeography, 28, 1351–1365, 2019.

Piedallu, C. and Gégout, J.-C.: Multiscale computation of solar radiation for predictive vegetation modelling, Annals of forest science, 64, 899–909, 2007.

Pollino, C. A., Woodberry, O., Nicholson, A., Korb, K., and Hart, B. T.: Parameterisation and evaluation of a Bayesian network for use in an ecological risk assessment, Environmental Modelling & Software, 22, 1140–1152, https://doi.org/10.1016/j.envsoft.2006.03.006, 2007.

Ramazi, P., Kunegel-Lion, M., Greiner, R., and Lewis, M. A.: Exploiting the full potential of Bayesian networks in predictive ecology, Methods in Ecology and Evolution, 12, 135–149, https://doi.org/10.1111/2041-210X.13509, 2021.

Reich, P. B. and Oleksyn, J.: Global patterns of plant leaf N and P in relation to temperature and latitude, Proceedings of the National Academy of Sciences, 101, 11001–11006, 2004.

Reichstein, M., Bahn, M., Mahecha, M. D., Kattge, J., and Baldocchi, D. D.: Linking plant and ecosystem functional biogeography, Proceedings of the National Academy of Sciences, 111, 13697–13702, https://doi.org/10.1073/pnas.1216065111, 2014.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat: Deep learning and process understanding for data-driven Earth system science, Nature, 566, 195–204, https://doi.org/10.1038/s41586-019-0912-1, 2019.

~~Rudin, C.: Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, Nat Mach Intell, 1, 206–215, https://doi.org/10.1038/s42256-019-0048-x, 2019.~~

Ryan, M. G. and Yoder, B. J.: Hydraulic limits to tree height and tree growth, Bioscience, 47, 235–242, 1997.

Sakschewski, B., von Bloh, W., Boit, A., Poorter, L., Peña-Claros, M., Heinke, J., Joshi, J., and Thonicke, K.: Resilience of Amazon forests emerges from plant trait diversity, Nature Clim Change, 6, 1032–1036, https://doi.org/10.1038/nclimate3109, 2016.

~~Santoro, M., Cartus, O., Carvalhais, N., Rozendaal, D. M. A., Avitabile, V., Araza, A., de Bruin, S., Herold, M., Quegan, S., Rodríguez-Veiga, P., Balzter, H., Carreiras, J., Schepaschenko, D., Korets, M., Shimada, M., Itoh, T., Moreno Martínez, Á., Cavlovic, J., Cazzolla Gatti, R., da Conceição Bispo, P., Dewnath, N., Labrière, N., Liang, J., Lindsell, J., Mitchard, E. T. A., Morel, A., Pacheco Pascagaza, A. M., Ryan, C. M., Slik, F., Vaglio Laurin, G., Verbeeck, H., Wijaya, A., and Willcock, S.: The global forest above-ground biomass pool for 2010 estimated from high-resolution satellite observations, Earth System Science Data, 13, 3927–3950, https://doi.org/10.5194/essd-13-3927-2021, 2021.~~

Santiago, L. S. and Mulkey, S. S.: Leaf productivity along a precipitation gradient in lowland Panama: patterns from leaf to ecosystem, Trees, 19, 349–356, https://doi.org/10.1007/s00468-004-0389-9, 2005.

Shi, H., Luo, G., Zheng, H., Chen, C., Bai, J., Liu, T., Ochege, F. U., and De Maeyer, P.: Coupling the water-energy-food-ecology nexus into a Bayesian network for water resources analysis and management in the Syr Darya River basin, Journal of Hydrology, 581, 124387, https://doi.org/10.1016/j.jhydrol.2019.124387, 2020a.

Shi, H., Luo, G., Zheng, H., Chen, C., Hellwich, O., Bai, J., Liu, T., Liu, S., Xue, J., Cai, P., He, H., Ochege, F. U., Van de Voorde, T., and de Maeyer, P.: A novel causal structure-based framework for comparing a basin-wide water–energy–food–ecology nexus applied to the data-limited Amu Darya and Syr Darya river basins, Hydrology and Earth System Sciences, 25, 901–925, https://doi.org/10.5194/hess-25-901-2021, 2021a.

Shi, H., Pan, Q., Luo, G., Hellwich, O., Chen, C., Voorde, T. V. de, Kurban, A., De Maeyer, P., and Wu, S.: Analysis of the Impacts of Environmental Factors on Rat Hole Density in the Northern Slope of the Tienshan Mountains with Satellite Remote Sensing Data, Remote Sensing, 13, 4709, https://doi.org/10.3390/rs13224709, 2021b.

Shi, H., Luo, G., Hellwich, O., Xie, M., Zhang, C., Zhang, Y., Wang, Y., Yuan, X., Ma, X., Zhang, W., Kurban, A., De Maeyer, P., and Van de Voorde, T.: Evaluation of water flux predictive models developed using eddy covariance observations and machine learning: a meta-analysis, Hydrology and Earth System Sciences Discussions, 1–21, https://doi.org/10.5194/hess-2022-90, 2022a.

Shi, H., Luo, G., Hellwich, O., Xie, M., Zhang, C., Zhang, Y., Wang, Y., Yuan, X., Ma, X., Zhang, W., Kurban, A., De Maeyer, P., and Van de Voorde, T.: Variability and uncertainty in flux-site-scale net ecosystem exchange simulations based on machine learning and remote sensing: a systematic evaluation, Biogeosciences, 19, 3739–3756, https://doi.org/10.5194/bg-19-3739-2022, 2022b.

Shi, Y., Jin, N., Ma, X., Wu, B., He, Q., Yue, C., and Yu, Q.: Attribution of climate and human activities to vegetation change in China using machine learning techniques, Agricultural and Forest Meteorology, 294, 108146, https://doi.org/10.1016/j.agrformet.2020.108146, 2020b.

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., and Zeileis, A.: Conditional variable importance for random forests, BMC Bioinformatics, 9, 307, https://doi.org/10.1186/1471-2105-9-307, 2008.

Toloşi, L. and Lengauer, T.: Classification with correlated features: unreliability of feature ranking and solutions, Bioinformatics, 27, 1986–1994, https://doi.org/10.1093/bioinformatics/btr300, 2011.

Tramontana, G., Jung, M., Schwalm, C. R., Ichii, K., Camps-Valls, G., Ráduly, B., Reichstein, M., Arain, M. A., Cescatti, A., Kiely, G., Merbold, L., Serrano-Ortiz, P., Sickert, S., Wolf, S., and Papale, D.: Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms, Biogeosciences, 13, 4291–4313, https://doi.org/10.5194/bg-13-4291-2016, 2016.

Trifonova, N., Kenny, A., Maxwell, D., Duplisea, D., Fernandes, J., and Tucker, A.: Spatio-temporal Bayesian network models with latent variables for revealing trophic dynamics and functional networks in fisheries ecology, Ecological Informatics, 30, 142–158, https://doi.org/10.1016/j.ecoinf.2015.10.003, 2015.

Wagner, F. H., Hérault, B., Rossi, V., Hilker, T., Maeda, E. E., Sanchez, A., Lyapustin, A. I., Galvão, L. S., Wang, Y., and Aragão, L. E.: Climate drivers of the Amazon forest greening, PLoS One, 12, e0180932, 2017.

Wang, Z., Zhu, D., Wang, X., Zhang, Y., and Peng, S.: Regressions underestimate the direct effect of soil moisture on land carbon sink variability, Global Change Biology, https://doi.org/10.1111/gcb.16422, 2022.

Weih, M. and Karlsson, P. S.: Growth response of Mountain birch to air and soil temperature: is increasing leaf-nitrogen content an acclimation to lower air temperature?, New Phytologist, 150, 147–155, https://doi.org/10.1046/j.1469-8137.2001.00078.x, 2001.

Wen, X.-F., Yu, G.-R., Sun, X.-M., Li, Q.-K., Liu, Y.-F., Zhang, L.-M., Ren, C.-Y., Fu, Y.-L., and Li, Z.-Q.: Soil moisture effect on the temperature dependence of ecosystem respiration in a subtropical Pinus plantation of southeastern China, Agricultural and Forest Meteorology, 137, 166–175, https://doi.org/10.1016/j.agrformet.2006.02.005, 2006.

Wever, L. A., Flanagan, L. B., and Carlson, P. J.: Seasonal and interannual variation in evapotranspiration, energy balance and surface conductance in a northern temperate grassland, Agricultural and Forest Meteorology, 112, 31–49, https://doi.org/10.1016/S0168-1923(02)00041-2, 2002.

Wright, I. J. and Westoby, M.: Leaves at low versus high rainfall: coordination of structure, lifespan and physiology, New phytologist, 155, 403–416, 2002.

Xu, L., Baldocchi, D. D., and Tang, J.: How soil moisture, rain pulses, and growth alter the response of ecosystem respiration to temperature, Global Biogeochemical Cycles, 18, 2004.

Xu, S., McVicar, T. R., Li, L., Yu, Z., Jiang, P., Zhang, Y., Ban, Z., Xing, W., Dong, N., Zhang, H., and Zhang, M.: Globally assessing the hysteresis between sub-diurnal actual evaporation and vapor pressure deficit at the ecosystem scale: Patterns and mechanisms, Agricultural and Forest Meteorology, 323, 109085, https://doi.org/10.1016/j.agrformet.2022.109085, 2022.

Yuan, W., Zheng, Y., Piao, S., Ciais, P., Lombardozzi, D., Wang, Y., Ryu, Y., Chen, G., Dong, W., Hu, Z., Jain, A. K., Jiang, C., Kato, E., Li, S., Lienert, S., Liu, S., Nabel, J. E. M. S., Qin, Z., Quine, T., Sitch, S., Smith, W. K., Wang, F., Wu, C., Xiao, Z., and Yang, S.: Increased atmospheric vapor pressure deficit reduces global vegetation growth, Science Advances, 5, eaax1396, https://doi.org/10.1126/sciadv.aax1396, 2019.

Zhou, S., Yu, B., Huang, Y., and Wang, G.: The effect of vapor pressure deficit on water use efficiency at the subdaily time scale, Geophysical Research Letters, 41, 5005–5013, https://doi.org/10.1002/2014GL060741, 2014.

Zhou, S., Williams, A. P., Berg, A. M., Cook, B. I., Zhang, Y., Hagemann, S., Lorenz, R., Seneviratne, S. I., and Gentine, P.: Land–atmosphere feedbacks exacerbate concurrent soil drought and atmospheric aridity, Proceedings of the National Academy of Sciences, 116, 18848–18853, 2019.