

1 **Revisiting and attributing the global controls on terrestrial**
2 **ecosystem functions of climate and plant traits at FLUXNET**
3 **sites via causal graphical models**

4 Haiyang Shi^{1,6}, Geping Luo^{2,3,4,6}, Olaf Hellwich⁷, Alishir Kurban^{2,3,4,6}, Philippe De Maeyer^{2,3,5,6} and Tim Van de
5 Voorde^{5,6}

6

7 ¹School of Earth Sciences and Engineering, Hohai University, Nanjing 211100, China.

8 ²State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese
9 Academy of Sciences, Urumqi, Xinjiang, 830011, China.

10 ³College of Resources and Environment, University of the Chinese Academy of Sciences, 19 (A) Yuquan Road,
11 Beijing, 100049, China.

12 ⁴The National Key Laboratory of Ecological Security and Sustainable Development in Arid Region (proposed),
13 Chinese Academy of Sciences, Urumqi, China.

14 ⁵Department of Geography, Ghent University, Ghent 9000, Belgium.

15 ⁶Sino-Belgian Joint Laboratory of Geo-Information, Ghent, Belgium.

16 ⁷Department of Computer Vision & Remote Sensing, Technische Universität Berlin, 10587 Berlin, Germany.

17

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

19 **Submitted to:** *Biogeosciences*

20

21 **Abstract**

22 Using statistical methods that not directly representing the causality between variables to attribute climate and
23 plant traits to control ecosystem function may lead to biased perceptions. We revisited this issue using a causal
24 graphical model, the Bayesian network (BN), capable of quantifying causality by conditional probability tables.
25 Based on expert knowledge and climate, vegetation, and ecosystem function data from the FLUXNET flux
26 stations, we constructed a BN representing the causal relationship of 'climate-plant trait-ecosystem function'.
27 Based on the sensitivity analysis function of the BN, we attributed the controls of climate and plant traits to
28 ecosystem function and compared the results with those based on Random forests and correlation analysis. The
29 main conclusions of this study include: BN can be used for the quantification of causal relationships between
30 complex ecosystems in response to climate change and enables the analysis of indirect effects among variables.
31 The causality reflected in the BN is as good as the expert knowledge of the causal links. Compared to BN, the
32 feature importance difference between 'mean vapor pressure deficit and cumulative soil water index' and
33 'maximum leaf area index and maximum vegetation height' reported by Random forests is higher and can be
34 overestimated. With the causality relation between correlated variables constructed, BN-based sensitivity
35 analysis can reduce the uncertainty in quantifying the importance of correlated variables. The understanding of
36 the mechanism of indirect effects of climate variables on ecosystem function through plant traits can be
37 deepened by the chain causality quantification in BNs.

38 **1 Introduction**

39 Ecosystem function is the capacity of natural processes and components to provide goods and services that
40 satisfy human needs, either directly or indirectly (de Groot et al., 2002). Ecosystem functions include the
41 physicochemical and biological processes within the ecosystem to maintain terrestrial life. Terrestrial
42 ecosystems have provided a variety of important ecosystem functions for our society (Manning et al., 2018).
43 Plant traits' role as important determinants of ecosystem functions has been widely recognized (Chapin III et al.,
44 2000), and various trait syndromes can result in distinct broad differences in ecosystem functions (Reichstein et
45 al., 2014). In the context of global climate change, it is also essential to understand the potential changes in
46 ecosystem functions (Grimm et al., 2013). The response of terrestrial ecosystem function to changes in climate,
47 plant traits, and the corresponding mechanisms, are complex due to enormous spatial and temporal variations
48 across ecosystems, climate zones, and also space-time scales (Diaz and Cabido, 1997; Madani et al., 2018;
49 Myers-Smith et al., 2019). Given the enormous variations, on the global scale, these issues have not been
50 clarified well.

51
52 In the past decades, measurements of ecosystem functions have been increasingly available to support studies of
53 the relations between ecosystem functions and climate variables. For example, eddy-covariance flux tower
54 observations (Baldocchi, 2014) for carbon flux (i.e., net ecosystem exchange (NEE)) and water flux (i.e.,
55 evapotranspiration (ET)) have been widely used to investigate changes in ecosystem functions and their
56 responses to climate change, vegetation condition changes, etc (Jung et al., 2020, 2010; Migliavacca et al., 2021;
57 Peaucelle et al., 2019). With the increase in such observations, various statistical analysis approaches such as
58 machine learning (Barnes et al., 2021; Migliavacca et al., 2021; Reichstein et al., 2019; Shi et al., 2022b, a;

59 Tramontana et al., 2016) have been used to mine the hidden information on the effects of climate change and its
60 induced changes in vegetation, etc. on ecosystem function variables such as carbon and water flux, which has
61 not been understood in depth by process-based models (e.g., biogeochemistry models (Sakschewski et al.,
62 2016)). For example, using Random forests (RF) and principal component analysis (PCA), a recent study
63 (Migliavacca et al., 2021) quantified the three main axes of terrestrial ecosystem function and their drivers based
64 on observations of carbon and water fluxes of FLUXNET stations (Pastorello et al., 2020) and various climate
65 and plant trait variables. Generally, data-driven approaches have become increasingly important recently in this
66 area (Reichstein et al., 2019).

67

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

83

84 Bayesian network (BN) is a causal graphical model based on conditional probability representation (Friedman et
85 al., 1997; Pearl, 1985) that characterizes the transmission of cause and effect through conditional probabilities
86 between variables. Currently, BN has been used in modeling causal relationships in many fields and has
87 demonstrated advantages in causal interpretation, including in the fields such as hydrology and ecology (Chan et
88 al., 2010; Keshtkar et al., 2013; Milns et al., 2010; Pollino et al., 2007; Shi et al., 2021a, b; Trifonova et al.,
89 2015). However, BN has rarely been used in the study of the attribution of changes in ecosystem function.
90 Therefore, this study used BN to attribute the controls of climate and plant traits to ecosystem function by
91 quantifying the causal relationships involved. The data used was from a previous study (Migliavacca et al.,
92 2021) which extracted ecosystem function, climate, and plant trait variables for FLUXNET flux stations. The
93 construction of the causal structure of BN referred to the previous expert knowledge of this system (Reichstein
94 et al., 2014). Further, by comparing BN-based attribution analysis, linear correlation analysis, and RF-based
95 IMP reported by the previous study (Migliavacca et al., 2021), we investigated the adding-values of using BN
96 for causal analysis and discussed its prospects in this paper.

97 **2 Methodology**98 **2.1 Data**

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

102 a) Ecosystem function variables: underlying Water Use Efficiency (uWUE), maximum evapotranspiration
103 (ETmax), maximum surface conductance (GSmax), maximum net CO₂ uptake of the ecosystem
104 (NEPmax), Gross Primary Productivity at light saturation (GPPsat), Mean basal ecosystem respiration at a
105 reference temperature of 15 °C (Rb), and apparent carbon-use efficiency (aCUE).

106 b) Plant trait variables: ecosystem scale foliar nitrogen concentration (Nmass), Maximum Leaf Area Index
107 (LAI_{max}), Maximum vegetation height (Hc). Of the total 202 sites (Migliavacca and Musavi, 2021), 101
108 sites have Nmass data, 153 sites have LAI_{max} data, and 199 sites have Hc data. Only 98 have data on all
109 these three plant trait variables.

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

112
113 These data have different producing processes, including those calculated from flux data, site records, extracted
114 from remote sensing data, etc. The detailed calculation methods can be found in Migliavacca et al., 2021.

115

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

Variable node	Definition and units	Type	Approach (Migliavacca et al., 2021)	Discretization in BN (equal quantile thresholds: 0%, 33.33%, 66.67%, and 100% percentile values)
uWUE	underlying Water Use Efficiency [gC kPa ^{0.5} kgH ₂ O ⁻¹]	Ecosystem function	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.	0.068, 2.51, 3.18, 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.	0.059, 0.17, 0.23, 0.423
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, 0.0077, 0.0123, 0.0566

NEPmax	maximum net CO ₂ 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).	1.953, 15.3, 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.	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, 2.07, 3.12, 10.67
aCUE	apparent carbon-use efficiency	Ecosystem function	aCUE was calculated by aCUE = 1- (Rb/GPP) and the median value of daily aCUE is used.	-1.19, 0.4, 0.74, 1
Nmass	ecosystem scale foliar nitrogen concentration [gN 100 g ⁻¹]	Plant trait	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).	0.65, 1.15, 1.76, 4.44
LAImax	Maximum Leaf Area Index [m ² m ⁻²]	Plant trait	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, 12.9
Hc	Maximum vegetation height [m]	Plant trait	Hc was collected from the literature (Migliavacca et al., 2011; Flechard et al., 2020), the BADM product, and/or site principal investigators.	0.04, 1.7, 16.0, 80.1
SWin	Mean incoming shortwave radiation [W m ⁻²]	Climate	SWin was from FLUXNET data.	54.43, 134.18, 182.44, 266.04
Tair	Mean temperature [degree C]	Climate	Tair was from FLUXNET data.	-10.45, 6.62, 14.73, 28.1
VPD	Mean Vapor Pressure Deficit [hPa]	Climate	VPD was from FLUXNET data.	0.62, 3.38, 5.76, 26.08
P	Mean annual precipitation [cm/year]	Climate	P was from FLUXNET data.	5.51, 45.28, 79.29, 256.61

CSWI	cumulative soil water index	Climate-related soil water availability	CSWI was computed as a measure of water availability (Nelson et al., 2018).	-93.49, -1.24, 2.01, 4.47
------	-----------------------------	---	---	---------------------------

117

118 **2.2 BN for analyzing causal relations**

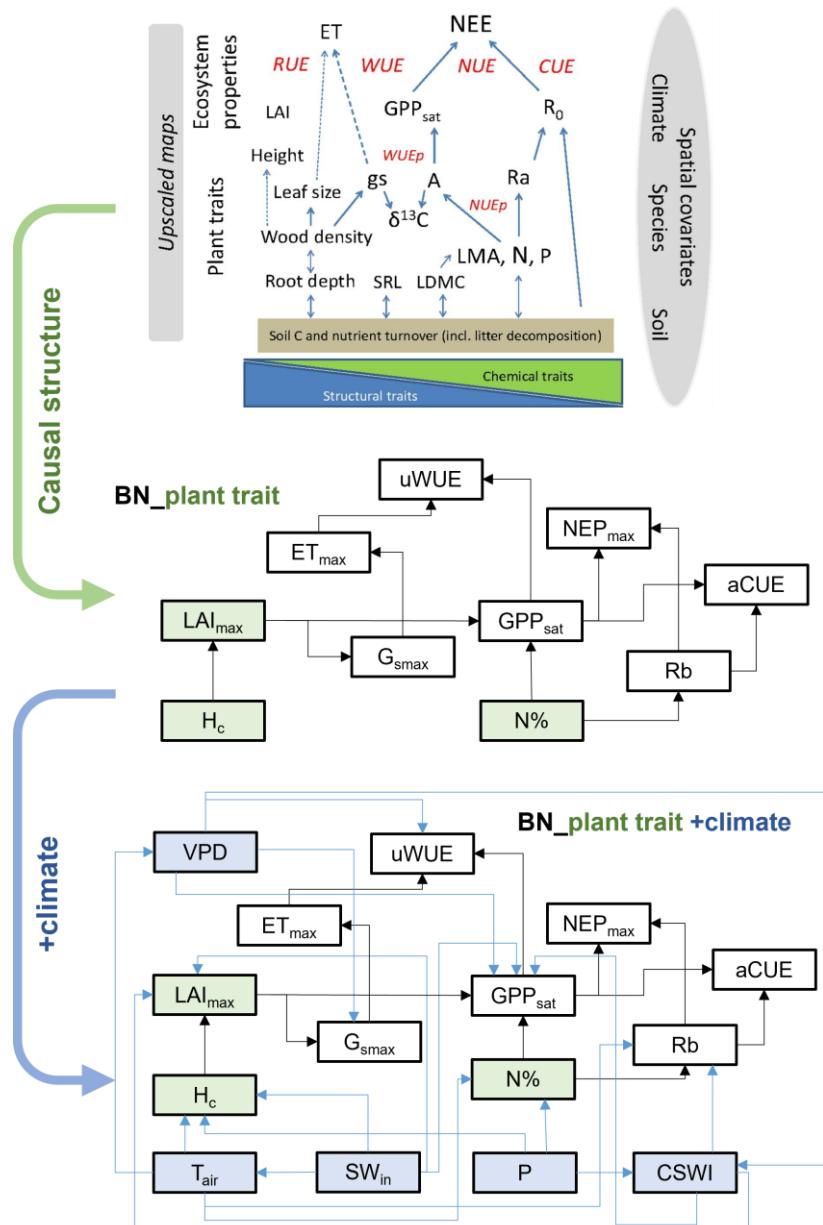
119 **2.2.1 BN structures**

120 Based on expert knowledge (Reichstein et al., 2014), we constructed the structure of BN containing the causal
 121 relationships between plant traits and ecosystem function variables: 'BN_plant_trait'. The causal links between
 122 the variables were referred to the relationship diagram in the upper part of Figure 1. Further, we added the
 123 climate variables and the corresponding causal relationships, expanding 'BN_plant_trait' to
 124 'BN_plant_trait_climate', which further incorporates the climate variables and their impacts on the system
 125 (Figure 1). The explanation of added causal links was shown in Table 2.

126

127 Each node is discretized for the BN compiling by the software Netica. The equal quantile (Nojavan A. et al.,
 128 2017) three-level discretization (the distribution of nodes (Figure S1) is divided into three levels) for each node
 129 is applied by the discretization thresholds of 0%, 33.33%, 66.67%, and 100% percentile values of the data
 130 distribution (Table 1) given the limitation of the amount of training data.

Expert Knowledge - Reichstein et al., 2014



131

132 Figure 1. The structure of two Bayesian networks (BNs) for attribution of variations in ecosystem functions.

133 'BN_plant_trait' in the median part incorporated the causal effects of plant traits (box in slight green) on
 134 ecosystem functions (box in white) from expert knowledge as the relation diagram on the upper part (Reichstein
 135 et al., 2014). 'BN_plant_trait_climate' in the lower part further incorporated the causal impacts of climate
 136 variables (box in light blue).

137

138 Table 2. Explanation of the added causal links between climate variable nodes, plant trait nodes, and ecosystem
 139 function variable nodes in the BNs.

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)
VPD	CSWI	CSWI declines under an increase in VPD	(Nelson et al., 2018)
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)
P	CSWI	CSWI declines under a decrease in P	(Nelson et al., 2018)
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)

141 **2.2.2 BN evaluation and node sensitivity analysis**

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

144
$$Pa(X) = Pa(X_1, X_2, \dots, X_n) = \prod_{i=1}^n Pa(X_i | pa(X_i)) \quad (1)$$

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

147

148 We used k-fold cross-validation to verify the reliability of the BN. The k-fold approach has been widely used in
149 previous studies for the validation of BNs (Marcot, 2012). In this study, k is set as 10 as commonly used
150 (Marcot and Hanea, 2021). We choose ETmax, GPPsat, and NEPmax for cross-validation of accuracy, and the
151 predicted status (status with the highest probability bar value) of the nodes will be compared with the actual
152 status and the classification accuracy will be calculated. These three nodes are the main terminal nodes and
153 primary objectives of the BN and represent the main water and carbon-related ecosystem functions, respectively.

154 The accuracy of these three variables can largely reflect the overall performance of BN.

155

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

159
$$MI = H(Q) - H(Q|F) = \sum_q \sum_f P(q, f) \log_2 \left(\frac{P(q, f)}{P(q)P(f)} \right) \quad (2)$$

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

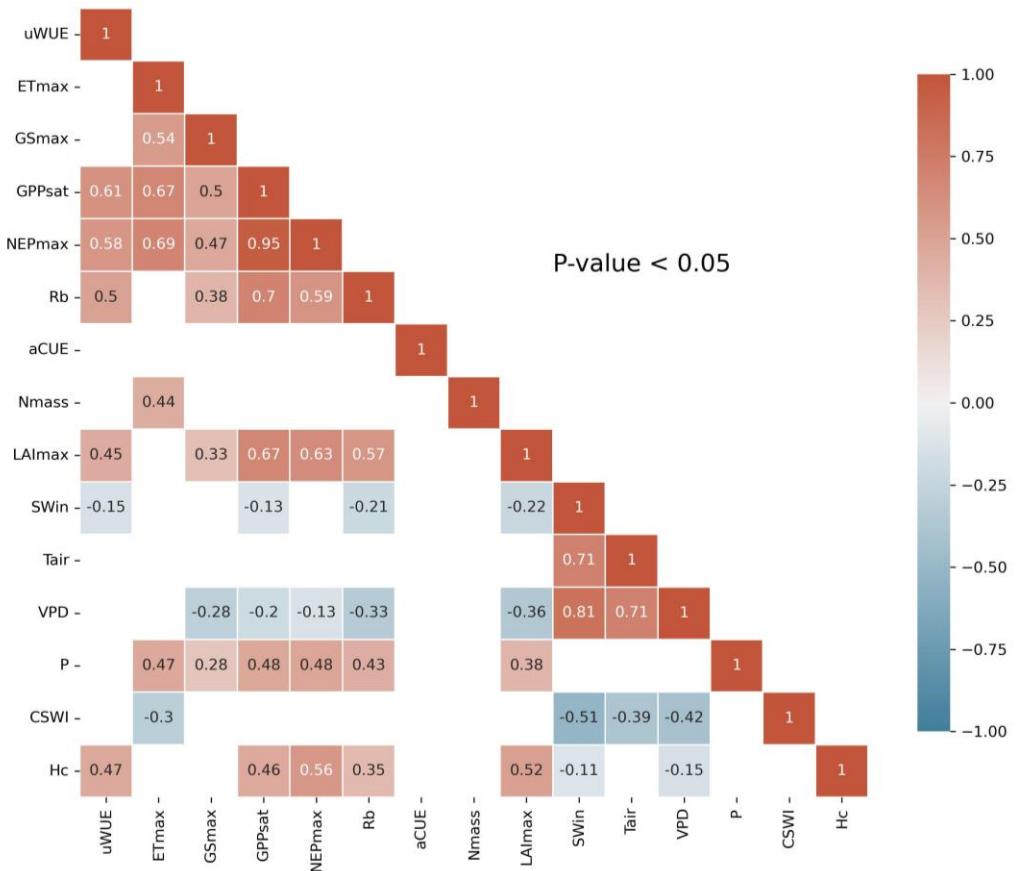
163 **2.2.3 Comparing different approaches used for attribution analysis**

164 Further, to clarify the adding-values of considering causality in the attribution analysis of controls on ecosystem
165 functions, the results of the BN-based sensitivity analysis (BN_sens) were compared with the other two
166 approaches. They are the results of the absolute values of additional linear correlation analysis (linear_corr) in
167 this study and the findings in Migliavacca et al., 2021 using RF feature importance (RF_imp). BN_sens and
168 linear_corr directly measure the effects of plant traits and climate variables on ecosystem function variables,
169 while RF_imp measures their effects on the three principal components (PC1, PC2, and PC3) of ecosystem
170 function variables, which were reported as the three major axes of ecosystem function by Migliavacca et al.,
171 2021. It was obtained from principal component analysis of 12 ecosystem function variables which included the
172 six variables uWUE, ETmax, GSmax, NEPmax, GPPsat, and Rb used in the methods BN_sens and linear_corr.
173 The first axis (PC1) explains 39.3% of the variance and is dominated by maximum ecosystem productivity
174 properties, as indicated by the loadings of GPPsat and NEPmax, and maximum evapotranspiration (ETmax).
175 The second axis (PC2) explains 21.4% of the variance and refers to water-use strategies as shown by the
176 loadings of water-use efficiency metrics, evaporative fraction, and GSmax. The third axis (PC3) explains 11.1%
177 of the variance and includes key attributes that reflect the carbon-use efficiency of ecosystems. PC3 is
178 dominated by apparent carbon-use efficiency, basal ecosystem respiration (Rb), and the amplitude of
179 evaporative fraction (Migliavacca et al., 2021).

180

181 **3 Results**182 **3.1 Correlation analysis**

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



189

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

192 **3.2 BN-based analysis**

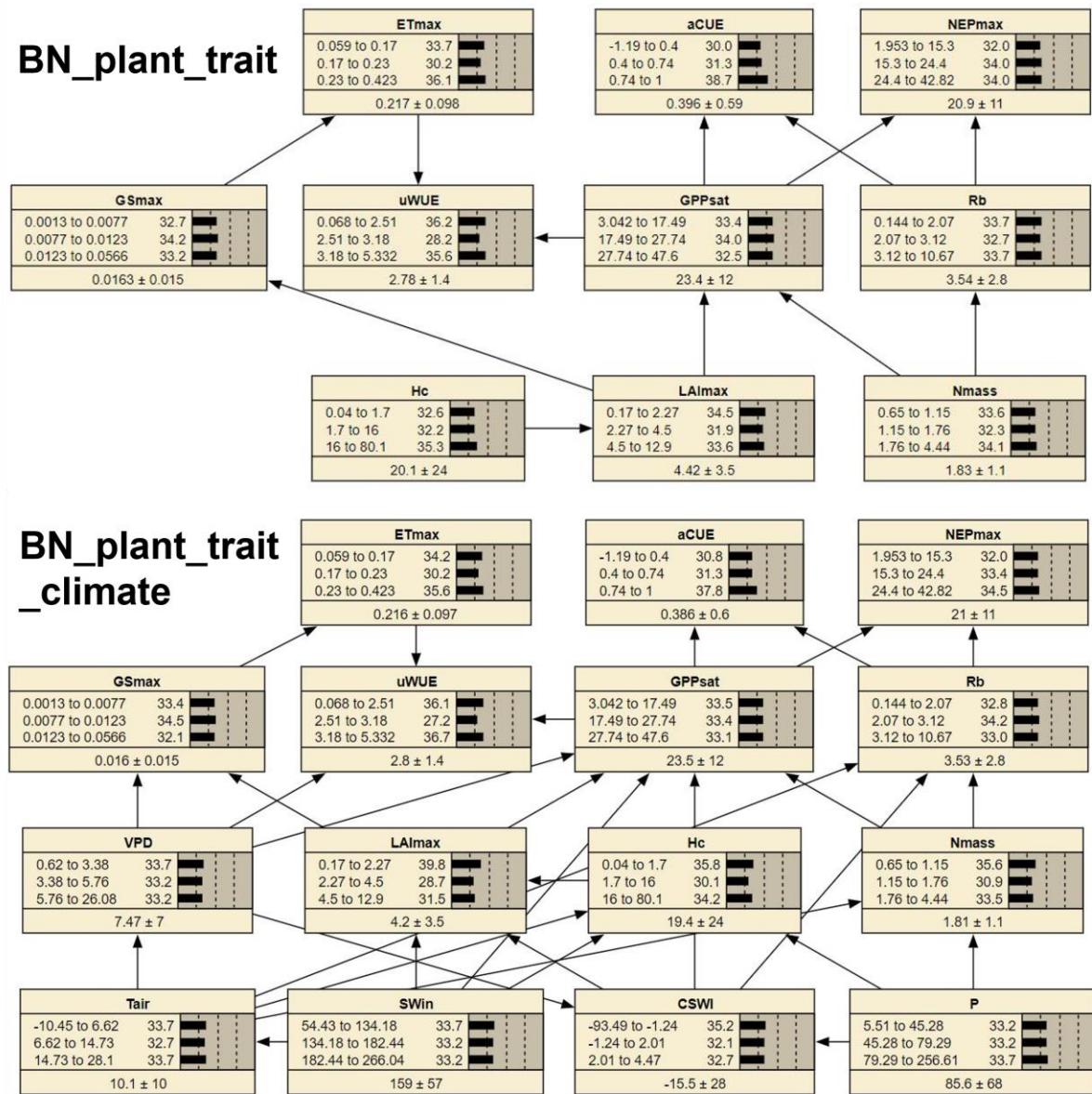
193 We compiled two different BNs (i.e., BN_plant_trait and BN_plant_trait_climate) (Figure 3) and found that the
 194 probability distributions of the values of the common nodes (ecosystem function and plant trait variable nodes)
 195 differed a little (e.g., in the probability distribution of LAImax, Hc, and Nmass) between the two BNs.
 196 Compared to BN_plant_trait, in BN_plant_trait_climate, the climate variables of sites with missing plant trait
 197 data forced the changes in the probability distributions of LAImax, Hc, and Nmass. In the EM algorithm, for

198 sites with missing plant trait data, existing relationships (obtained from observations from other sites) between
 199 plant trait variables and climate variables are used in the data interpolation of plant trait variables. In
 200 BN_plant_trait_climate, the added linkages of climate variables to plant trait variables resulted in higher
 201 probability values of the low-value status of the plant trait variables.

202

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

206



207

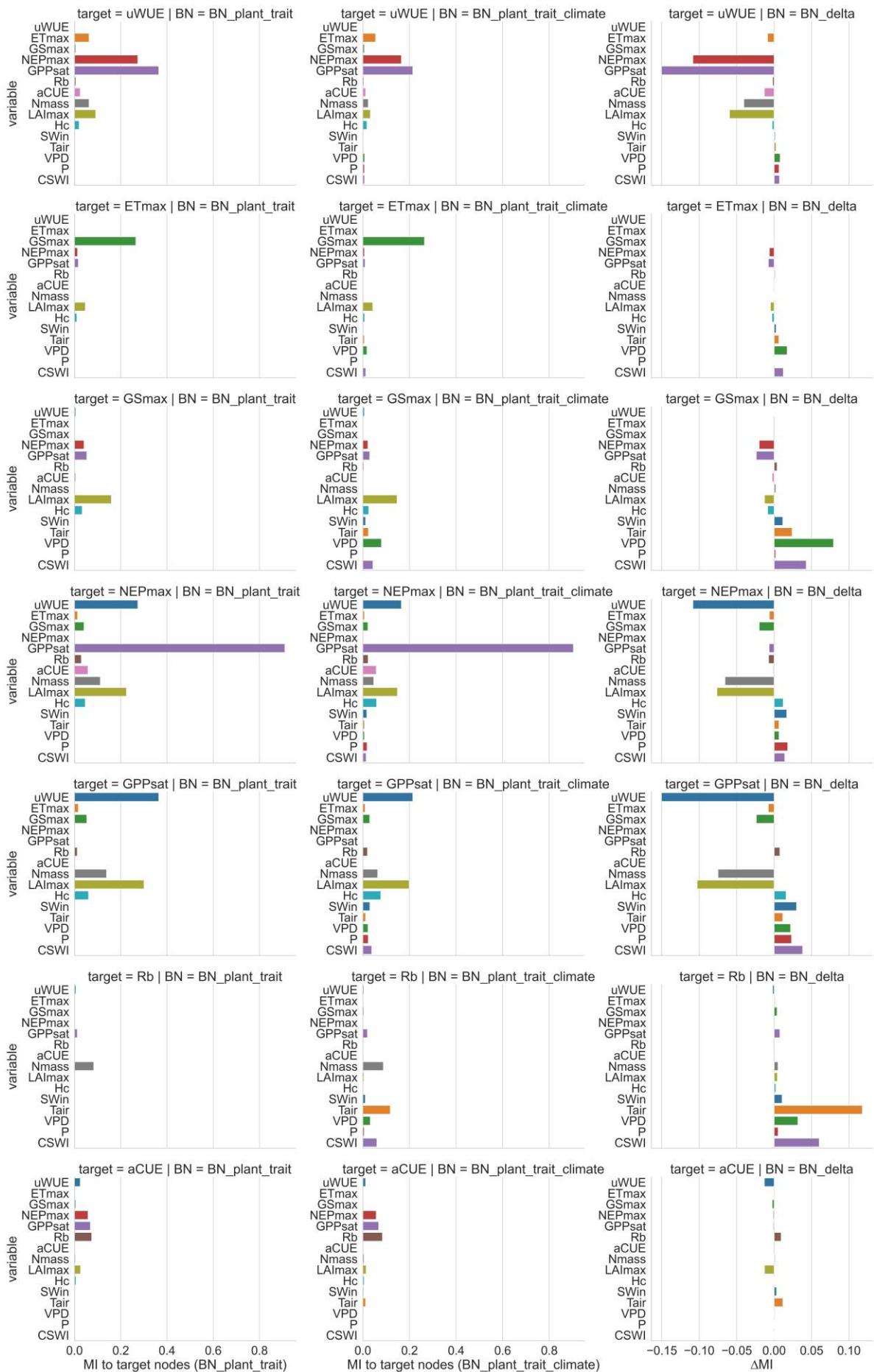
208 Figure 3. The compiled two BNs ('BN_plant_trait' and 'BN_plant_trait_climate'). The bars of each node
 209 represent its probability distribution. At the bottom part of each node, the left and right side values of the '±' are
 210 the mean and standard deviation of the distribution, respectively.

211

212 We performed sensitivity analyses (Figure 4) on the ecosystem function variables in both BNs to assess their
213 sensitivity to various climate and plant trait variables. We also calculated the difference in sensitivity MI
214 between the two BNs (Figure 4) to compare the change in sensitivity of ecosystem function to each variable
215 after adding further climate variables to the plant trait variables only. The sensitivity of different ecosystem
216 function variables to plant traits and climate variables was highly variable in both BNs. The magnitude of
217 sensitivity of ecosystem function nodes to plant traits and climate variables was related to whether these plant
218 traits and climate variables were set as their parent nodes. In BN_plant_trait, for the carbon fluxes GPPsat and
219 NEPmax, Nmass, and LAImax had higher sensitivity due to Nmass and LAI being set as their parent nodes. For
220 the water flux ETmax, it does not have high sensitivity to plant trait variables such as LAImax and Hc, although
221 these plant trait variables are set as the parent nodes of ETmax. This indicates the difference in the strength of
222 the control effects of plant traits on carbon and water fluxes.

223

224 In the sensitivity analysis of BN_plant_trait_climate, the sensitivity patterns of the ecosystem function variables
225 changed as a result of the inclusion of climate variables and the change in causality they introduced. The
226 sensitivity of the ecosystem function variables to climate variables was significantly increased (especially for
227 Tair, VPD, and CSWI). The control of plant traits on ecosystem function in BN_plant_trait is also partially
228 transformed into an indirect effect of climate variables by first controlling plant trait variables and then
229 controlling ecosystem function. For example, in BN_plant_trait_climate, for GPPsat, a decrease in the
230 sensitivity of GPPsat to LAImax and an increase in the sensitivity to Tair was observed after the causal chain of
231 Tair influencing Hc, LAImax, and then GPPsat was set. This can be explained by the fact that higher
232 temperatures promote vegetation growth and thus may increase LAImax, which then indirectly alters the
233 probability distribution of the GPPsat node. In previous studies based on statistical methods that did not consider
234 the chain causality, this indirect control on GPPsat from Tair may have been included in the contribution of
235 LAImax to GPPsat. Similarly, a chain causality of P by first affecting Nmass and then indirectly GPPsat was
236 also found. However, the effect of P by first affecting Hc, LAImax, and then indirectly affecting ETmax and
237 GSmax appears to be not large.



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

246 **3.3 Comparing results from RF-based, BN-based analysis, and correlation analysis**

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

257

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

	Methods	Nmass	LAmax	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								

264

265 Figure 5. Comparisons of relationships of ecosystem functional variables to plant traits and climate variables in
 266 different analyses. Method RF_imp is Random forest variable importance (Migliavacca et al., 2021) (see
 267 Methodology section). Method linear_corr is Linear correlation analysis with the absolute values of Pearson
 268 correlation coefficients (see Methodology section). Method BN_sens is a BN-based sensitivity analysis with
 269 sensitivity values MI reported. The values in each method group are in red for high values and in blue for low
 270 values. The color depth is dependent on values and the scale is the same in each row.

271 4 Discussions

272 Based on BN, this study investigates the prospect of using causal graphical models to revisit and attribute the
 273 control of climate and plant trait variations to ecosystem functions. Because of the inclusion of the constraints
 274 provided by expert knowledge (Reichstein et al., 2014) and other perceptions from many previous studies, BN-
 275 based attribution analysis is relatively reliable in terms of the represented mechanisms of causal links. It can
 276 update our knowledge of the contribution of some teleconnection variables through causal chains. The effective
 277 implementation of BN-based causal analysis may depend on the reliability of the causal relationships provided
 278 by expert knowledge (directional links between variables). We can establish the connection relationships and
 279 network structures between variables from expert knowledge and assign the specific quantification of the

280 connection relationships (conditional probability tables) to observations (Shi et al., 2021a). If further combined
281 with findings from process-based models, it is promising to significantly improve our understanding of the
282 complex ‘climate-plant trait-ecosystem function’ relationships by comparing detailed relationships and structural
283 influences between variables.

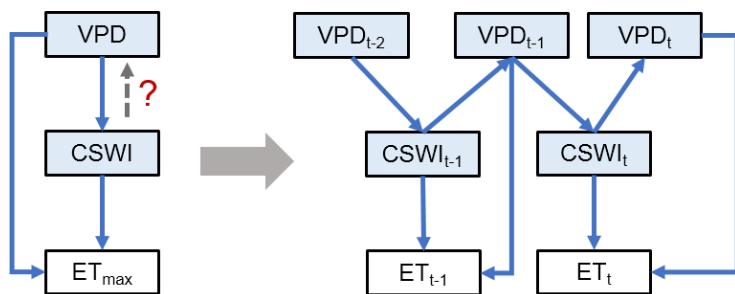
284

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

311

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

319 this correlation coefficient can be explored. Such implements may be useful to deepen the impact of various
 320 higher order effects.
 321
 322 Besides, the BN in this study was mainly based on data averaged over multiple years, thus possibly partially
 323 underestimating the effect of temporal variations in the relationships between variables. Another limitation of
 324 the BN proposed above is that the causal relationships between variables are unidirectional, while it is difficult
 325 to represent interactions and feedback between variables (Marcot and Penman, 2019). In future studies, to
 326 address these two issues, BN based on temporal dynamics can be promising (Figure 6). By refining the
 327 interaction of temporal lags between variables, it is possible to incorporate not only temporal variation but also
 328 control factors that attribute interactions and feedback between variables. For example, the interaction and
 329 feedback mechanisms of VPD, soil moisture, and ET with lag effects (Figure 6) and their impacts on ecosystems
 330 have attracted extensive interest from researchers (Anderegg et al., 2019; Humphrey et al., 2021; Lansu et al.,
 331 2020; Liu et al., 2020; Xu et al., 2022; Zhou et al., 2019), but conventional statistical methods have been
 332 ineffective in analyzing such relationships with both interactive causality and temporal lags. In contrast, the BN
 333 proposed here, which incorporates feedback effects and lagged effects that were common in climate-ecosystem
 334 relations (Lin et al., 2019), is potentially able to address this issue from a data-driven approach. In the practical
 335 modeling, different periods of the same node may still be not independent. Therefore, the split scheme of such
 336 periods may be critical. For example, a period between two precipitation events can be treated as one sample,
 337 which can enhance independence between periods. Subsequently, a such period can be divided into smaller
 338 periods such as t , $t-1$, $t-2$, etc. to aggregate the node values to appropriate time scales. Thus one sample can
 339 represent the interaction relationship between variables with lags in this period. Finally, we can integrate records
 340 of such periods between two precipitation events from sites across different climate zones and biomes to build
 341 synthesis models for global analysis of such problems. Such research frameworks in BN-based modeling may
 342 be difficult due to high computational costs given the large amount of data. Fortunately, recently proposed new
 343 causal models have the potential to address this limitation, such as the introduction of causality into deep
 344 learning frameworks (Luo et al., 2020; Cui and Athey, 2022). If further combined with the findings of process-
 345 based models, our understanding of climate and ecosystem interactions and feedback and their mechanisms in
 346 time is hopefully deepened.
 347



348
 349 Figure 6. The future BNs with the temporal causality further considered addressing the causality of the
 350 interaction between variables. The VPD-CSWI-ET relationship is used here as an example. t , $t-1$, and $t-2$ denote
 351 the current period, the last period, and the period before the last period, respectively. The network on the left
 352 only considers the effect of VPD on CSWI without considering the feedback of CSWI on the VPD. The network

353 on the right characterizes the VPD-CSWI interaction with the feedback from CSWI at period t-1 to VPD at
354 period t.

355 **5 Conclusion**

356 Based on BN, we revisited and attributed the contribution of climate and plant traits to global terrestrial
357 ecosystem function. The major conclusions of this study include:

- 358 1. BN can be used for the quantification of causal relationships between complex ecosystems in response to
359 climate change and enables the analysis of indirect effects among variables.
- 360 2. Compared to BN, the feature importance difference between 'VPD and CSWI' and 'LAI_{max} and H_c'
361 reported by Random forests is higher and can be overestimated.
- 362 3. With the causality relation between correlated variables constructed, BN_sens can reduce the uncertainty in
363 quantifying the importance of correlated variables.
- 364 4. The understanding of the mechanism of indirect effects of climate variables on ecosystem function through
365 plant traits can be deepened by the chain causality quantification in BNs.

366

367 **Acknowledgements**

368 We are grateful to the two anonymous reviewers for their thorough and careful review that led to substantial
369 improvements in the manuscript. We are also grateful to Dr. Kirsten Thonicke for being the associate editor.

370 **Financial support**

371 This research was supported by the Tianshan Talent Cultivation (Grant No. 2022TSYCLJ0001), the Key
372 projects of the Natural Science Foundation of Xinjiang Autonomous Region (Grant No. 2022D01D01), the
373 Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA20060302), and
374 High-End Foreign Experts Project.

375 **Author Contributions**

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

379 **Competing interests**

380 The authors declare that they have no conflict of interest.

381 **Code availability**

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

384 **Data availability**

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

387

388 **References**

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

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

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

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

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

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

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

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

411 Cui, P. and Athey, S.: Stable learning establishes some common ground between causal inference and
412 machine learning, *Nat Mach Intell*, 4, 110–115, <https://doi.org/10.1038/s42256-022-00445-z>, 2022.

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

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

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

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

422 Flechard, C. R., Ibrom, A., Skiba, U. M., de Vries, W., van Oijen, M., Cameron, D. R., Dise, N. B.,
423 Korhonen, J. F. J., Buchmann, N., Legout, A., Simpson, D., Sanz, M. J., Aubinet, M., Loustau, D.,
424 Montagnani, L., Neirynck, J., Janssens, I. A., Pihlatie, M., Kiese, R., Siemens, J., Francez, A.-J.,
425 Augustin, J., Varlagin, A., Olejnik, J., Juszczak, R., Aurela, M., Berveiller, D., Chojnicki, B. H.,
426 Dämmgen, U., Delpierre, N., Djuricic, V., Dreher, J., Dufrêne, E., Eugster, W., Fauvel, Y., Fowler,
427 D., Frumau, A., Granier, A., Gross, P., Hamon, Y., Helfter, C., Hensen, A., Horváth, L., Kitzler, B.,
428 Kruijt, B., Kutsch, W. L., Lobo-do-Vale, R., Lohila, A., Longdoz, B., Marek, M. V., Matteucci, G.,

429 Mitosinkova, M., Moreaux, V., Neftel, A., Ourcival, J.-M., Pilegaard, K., Pita, G., Sanz, F.,
430 Schjoerring, J. K., Sebastià, M.-T., Tang, Y. S., Uggerud, H., Urbaniak, M., van Dijk, N., Vesala, T.,
431 Vidic, S., Vincke, C., Weidinger, T., Zechmeister-Boltenstern, S., Butterbach-Bahl, K., Nemitz, E.,
432 and Sutton, M. A.: Carbon–nitrogen interactions in European forests and semi-natural vegetation –
433 Part 1: Fluxes and budgets of carbon, nitrogen and greenhouse gases from ecosystem monitoring and
434 modelling, *Biogeosciences*, 17, 1583–1620, <https://doi.org/10.5194/bg-17-1583-2020>, 2020.

435 Fleischer, K., Wårland, D., Van der Molen, M. K., Rebel, K. T., Arneth, A., Erisman, J. W., Wassen,
436 M. J., Smith, B., Gough, C. M., and Margolis, H. A.: Low historical nitrogen deposition effect on
437 carbon sequestration in the boreal zone, *Journal of Geophysical Research: Biogeosciences*, 120,
438 2542–2561, 2015.

439 Friedman, N., Geiger, D., and Goldszmidt, M.: Bayesian network classifiers, *Machine learning*, 29,
440 131–163, 1997.

441 Green, J. K., Seneviratne, S. I., Berg, A. M., Findell, K. L., Hagemann, S., Lawrence, D. M., and
442 Gentine, P.: Large influence of soil moisture on long-term terrestrial carbon uptake, *Nature*, 565, 476–
443 479, 2019.

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

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

450 Grossiord, C., Buckley, T. N., Cernusak, L. A., Novick, K. A., Poulter, B., Siegwolf, R. T. W.,
451 Sperry, J. S., and McDowell, N. G.: Plant responses to rising vapor pressure deficit, *New Phytologist*,
452 226, 1550–1566, <https://doi.org/10.1111/nph.16485>, 2020.

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

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

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

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

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

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

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

477 Koch, G. W., Sillett, S. C., Jennings, G. M., and Davis, S. D.: The limits to tree height, *Nature*, 428,
478 851–854, 2004.

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

481 Lansu, E. M., van Heerwaarden, C., Stegehuis, A. I., and Teuling, A. J.: Atmospheric aridity and
482 apparent soil moisture drought in European forest during heat waves, *Geophysical Research Letters*,
483 47, e2020GL087091, 2020.

484 Lin, C., Gentine, P., Frankenberg, C., Zhou, S., Kennedy, D., and Li, X.: Evaluation and mechanism
485 exploration of the diurnal hysteresis of ecosystem fluxes, *Agricultural and Forest Meteorology*, 278,
486 107642, <https://doi.org/10.1016/j.agrformet.2019.107642>, 2019.

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

489 Liu, Q., Fu, Y. H., Zeng, Z., Huang, M., Li, X., and Piao, S.: Temperature, precipitation, and
490 insolation effects on autumn vegetation phenology in temperate China, *Global Change Biology*, 22,
491 644–655, <https://doi.org/10.1111/gcb.13081>, 2016.

492 Luo, Y., Peng, J., and Ma, J.: When causal inference meets deep learning, *Nat Mach Intell*, 2, 426–
493 427, <https://doi.org/10.1038/s42256-020-0218-x>, 2020.

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

498 Manning, P., Van Der Plas, F., Soliveres, S., Allan, E., Maestre, F. T., Mace, G., Whittingham, M. J.,
499 and Fischer, M.: Redefining ecosystem multifunctionality, *Nature ecology & evolution*, 2, 427–436,
500 2018.

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

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

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

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

510 Migliavacca, M., Reichstein, M., Richardson, A. D., Colombo, R., Sutton, M. A., Lasslop, G.,
511 Tomelleri, E., Wohlfahrt, G., Carvalhais, N., and Cescatti, A.: Semiempirical modeling of abiotic and

512 biotic factors controlling ecosystem respiration across eddy covariance sites, *Global Change Biology*,
513 17, 390–409, 2011.

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

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

527 Moles, A. T., Warton, D. I., Warman, L., Swenson, N. G., Laffan, S. W., Zanne, A. E., Pitman, A.,
528 Hemmings, F. A., and Leishman, M. R.: Global patterns in plant height, *Journal of ecology*, 97, 923–
529 932, 2009.

530 Monteith, J. L.: Solar radiation and productivity in tropical ecosystems, *Journal of applied ecology*, 9,
531 747–766, 1972.

532 Moon, T. K.: The expectation-maximization algorithm, *IEEE Signal processing magazine*, 13, 47–60,
533 1996.

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

538 Musavi, T., Migliavacca, M., van de Weg, M. J., Kattge, J., Wohlfahrt, G., van Bodegom, P. M.,
539 Reichstein, M., Bahn, M., Carrara, A., and Domingues, T. F.: Potential and limitations of inferring
540 ecosystem photosynthetic capacity from leaf functional traits, *Ecology and evolution*, 6, 7352–7366,
541 2016.

542 Myers-Smith, I. H., Thomas, H. J. D., and Bjorkman, A. D.: Plant traits inform predictions of tundra
543 responses to global change, *New Phytologist*, 221, 1742–1748, <https://doi.org/10.1111/nph.15592>,
544 2019.

545 Nelson, J. A., Carvalhais, N., Migliavacca, M., Reichstein, M., and Jung, M.: Water-stress-induced
546 breakdown of carbon–water relations: indicators from diurnal FLUXNET patterns, *Biogeosciences*,
547 15, 2433–2447, 2018.

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

551 Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W., Poindexter, C., Chen,
552 J., Elbashandy, A., Humphrey, M., Isaac, P., Polidori, D., Reichstein, M., Ribeca, A., van Ingen, C.,
553 Vuichard, N., Zhang, L., Amiro, B., Ammann, C., Arain, M. A., Ardö, J., Arkebauer, T., Arndt, S. K.,
554 Arriga, N., Aubinet, M., Aurela, M., Baldocchi, D., Barr, A., Beamesderfer, E., Marchesini, L. B.,
555 Bergeron, O., Beringer, J., Bernhofer, C., Berveiller, D., Billesbach, D., Black, T. A., Blanken, P. D.,

556 Bohrer, G., Boike, J., Bolstad, P. V., Bonal, D., Bonnefond, J.-M., Bowling, D. R., Bracho, R.,
557 Brodeur, J., Brümmer, C., Buchmann, N., Burban, B., Burns, S. P., Buysse, P., Cale, P., Cavagna, M.,
558 Cellier, P., Chen, S., Chini, I., Christensen, T. R., Cleverly, J., Collalti, A., Consalvo, C., Cook, B. D.,
559 Cook, D., Coursolle, C., Cremonese, E., Curtis, P. S., D'Andrea, E., da Rocha, H., Dai, X., Davis, K.
560 J., Cinti, B. D., Grandcourt, A. de, Ligne, A. D., De Oliveira, R. C., Delpierre, N., Desai, A. R., Di
561 Bella, C. M., Tommasi, P. di, Dolman, H., Domingo, F., Dong, G., Dore, S., Duce, P., Dufrêne, E.,
562 Dunn, A., Dušek, J., Eamus, D., Eichelmann, U., ElKhidir, H. A. M., Eugster, W., Ewenz, C. M.,
563 Ewers, B., Famulari, D., Fares, S., Feigenwinter, I., Feitz, A., Fensholt, R., Filippa, G., Fischer, M.,
564 Frank, J., Galvagno, M., et al.: The FLUXNET2015 dataset and the ONEFlux processing pipeline for
565 eddy covariance data, *Sci Data*, 7, 225, <https://doi.org/10.1038/s41597-020-0534-3>, 2020.

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

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

572 Peaucelle, M., Bacour, C., Ciais, P., Vuichard, N., Kuppel, S., Peñuelas, J., Belotti Marchesini, L.,
573 Blanken, P. D., Buchmann, N., and Chen, J.: Covariations between plant functional traits emerge from
574 constraining parameterization of a terrestrial biosphere model, *Global ecology and biogeography*, 28,
575 1351–1365, 2019.

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

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

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

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

586 Reichstein, M., Bahn, M., Mahecha, M. D., Kattge, J., and Baldocchi, D. D.: Linking plant and
587 ecosystem functional biogeography, *Proceedings of the National Academy of Sciences*, 111, 13697–
588 13702, <https://doi.org/10.1073/pnas.1216065111>, 2014.

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

592 Ryan, M. G. and Yoder, B. J.: Hydraulic limits to tree height and tree growth, *Bioscience*, 47, 235–
593 242, 1997.

594 Sakschewski, B., von Bloh, W., Boit, A., Poorter, L., Peña-Claros, M., Heinke, J., Joshi, J., and
595 Thonicke, K.: Resilience of Amazon forests emerges from plant trait diversity, *Nature Clim Change*,
596 6, 1032–1036, <https://doi.org/10.1038/nclimate3109>, 2016.

597 Santiago, L. S. and Mulkey, S. S.: Leaf productivity along a precipitation gradient in lowland Panama:
598 patterns from leaf to ecosystem, *Trees*, 19, 349–356, <https://doi.org/10.1007/s00468-004-0389-9>,
599 2005.

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

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

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

613 Shi, H., Luo, G., Hellwich, O., Xie, M., Zhang, C., Zhang, Y., Wang, Y., Yuan, X., Ma, X., Zhang,
614 W., Kurban, A., De Maeyer, P., and Van de Voorde, T.: Evaluation of water flux predictive models
615 developed using eddy-covariance observations and machine learning: a meta-analysis, *Hydrology and*
616 *Earth System Sciences*, 26, 4603–4618, <https://doi.org/10.5194/hess-26-4603-2022>, 2022a.

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

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

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

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

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

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

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

639 Weih, M. and Karlsson, P. S.: Growth response of Mountain birch to air and soil temperature: is
640 increasing leaf-nitrogen content an acclimation to lower air temperature?, *New Phytologist*, 150, 147–
641 155, <https://doi.org/10.1046/j.1469-8137.2001.00078.x>, 2001.

642 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,
643 Z.-Q.: Soil moisture effect on the temperature dependence of ecosystem respiration in a subtropical
644 Pinus plantation of southeastern China, *Agricultural and Forest Meteorology*, 137, 166–175,
645 <https://doi.org/10.1016/j.agrformet.2006.02.005>, 2006.

646 Wever, L. A., Flanagan, L. B., and Carlson, P. J.: Seasonal and interannual variation in
647 evapotranspiration, energy balance and surface conductance in a northern temperate grassland,
648 *Agricultural and Forest Meteorology*, 112, 31–49, [https://doi.org/10.1016/S0168-1923\(02\)00041-2](https://doi.org/10.1016/S0168-1923(02)00041-2),
649 2002.

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

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

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

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

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

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

669