



HAL
open science

Assessing dependence between soil ecosystem services as a function of weather and soil: Application of vine copula modeling

Tristan Senga Kiessé, Blandine Lemerrier, Michael S. Corson, Yosra Ellili-Bargaoui, Jihad Afassi, Christian Walter

► To cite this version:

Tristan Senga Kiessé, Blandine Lemerrier, Michael S. Corson, Yosra Ellili-Bargaoui, Jihad Afassi, et al.. Assessing dependence between soil ecosystem services as a function of weather and soil: Application of vine copula modeling. *Environmental Modelling and Software*, 2024, 172, pp.105920. 10.1016/j.envsoft.2023.105920 . hal-04326968

HAL Id: hal-04326968

<https://hal.inrae.fr/hal-04326968v1>

Submitted on 23 Aug 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

1 **Assessing dependence between soil ecosystem services as a function of weather and soil:**
2 **application of vine copula modeling**

3 Tristan Senga Kiessé^{1,*}, Blandine Lemerrier¹, Michael S. Corson¹, Yosra Ellili-Bargaoui²,
4 Jihad Afassi¹, Christian Walter¹

5 ¹INRAE, Institut Agro, SAS, F-35000 Rennes, France

6 ²Interact, UniLaSalle, F-60000 Beauvais, France

7 *Corresponding author: tristan.senga-kiesse@inrae.fr ORCID iD: [0000-0003-2710-5825](https://orcid.org/0000-0003-2710-5825)

8 **Abstract**

9 Soils are natural ecosystems that provide ecosystem services, whose provision depends on
10 multiple soil properties, climate conditions and human management. Dependence among soil
11 ecosystem services (SESs) must therefore be considered to reliably assess risks of improving
12 SES, as a function of weather conditions or soil properties. The present study described
13 dependence among regulating and provisioning SESs predicted by a biophysical soil and crop
14 model, based on a dataset of soils in France. We applied vine copula modeling as a statistical
15 method that can model joint distribution functions of three SESs and enabled us to estimate
16 probabilities of exceeding a level of one SES as a function of another SES. Trade-offs may need
17 to be made between them to manage soil and water resources and achieve a given yield. By
18 highlighting the degree of dependence among multiple SESs, copula models thus provide
19 information that may improve understanding or management of ESs.

20 **Keywords:** dependence; soil ecosystem services; soil properties; weather conditions

21 **Highlights**

- 22 • Copulas modeled variable dependence between soil ecosystem services (SESs)
23 • Vine copula models analyzed multiple dependence by using pair-decomposition of SESs
24 • One SES could affect two other SESs separately, but not the dependence between them
25

26 **1. Introduction**

27 Soils are the focus of many studies since their multifunctional role is crucial as the central
28 interface among the lithosphere, biosphere, hydrosphere and atmosphere (Dominati et al.,
29 2014). Soils contribute to human welfare through the ecosystem services (ESs) that they provide

30 (Ellili-Bargaoui et al., 2021). Soil ESs (SESs) can be assessed using multiple indicators
31 obtained by modeling biophysical processes and measuring multiple soil properties. For
32 instance, maximum rooting depth and soil available water capacity are the indicators frequently
33 used for ESs such as groundwater storage, plant biomass provision and water-quality regulation.
34 Because ESs depend on the same biophysical processes and soil properties, many studies focus
35 on dependence among ES to help develop policy strategies and make decisions about natural-
36 resource management (Nelson et al., 2009). Describing dependence among ES highlights
37 positive or negative correlations in which increasing one ES increases or decreases another one
38 (i.e., a synergy or trade-off, respectively) (Lee and Lautenbach, 2016). Thus, there is practical
39 interest in approaches that can investigate the distribution of ES of interest given the values of
40 other ES, as a function of the environmental context (e.g., weather conditions, soil properties).

41 A variety of methods have been used to analyze relationships among ESs, such as regression
42 models (e.g., generalized linear, additive), statistical tests of categorical variables (e.g., chi-
43 square), principal component analysis (PCA) of quantitative variables and pairwise correlation
44 coefficients, which is one of the most popular quantitative methods (Mouchet et al., 2014). PCA
45 can help highlight contrasting relationships among SES and identify groups of soil sites as a
46 function of given soil characteristics. Structural equation modeling, which is based on
47 regression models that relate measured variables to “latent” (i.e., unmeasured) variables, can
48 also help understand causal relationships between ES and explanatory variables (Grace, 2006;
49 Leong et al., 2020). Relationships among ESs have been visualized using radial diagrams
50 (Calzolari et al., 2016) and bagplots (Ellili-Bargaoui et al., 2021), and the latter have been used
51 to map ESs in order to manage the ESs later (Egoh et al., 2008). In addition, optimal trade-offs
52 between two SESs can be studied by plotting a curve (the “production possibility frontier” (Wu
53 et al., 2021)) based on estimated responses of SESs to management practices in a variety of
54 scenarios (see also Joly et al. (2021)). However, these approaches are generally limited by
55 assumptions about input variables (e.g., independence), underlying data (e.g., normal
56 distribution) or the shape (e.g., linear) of relationships between input variables and output
57 variables used in models.

58 Copula modeling is a statistical method useful for constructing joint distribution functions that
59 describe dependence between random variables (Sklar, 1959; Joe, 1993). They do not require
60 making any assumptions about variables (e.g., that they follow a normal distribution) (Ching et
61 al., 2014), unlike other statistical methods, and are particularly useful for detecting dependence
62 when variables have simultaneously high or low values (i.e., “upper” or “lower tail-

63 dependence”, respectively) (Coles et al., 1999; Embrechts et al., 2002), which can occur during
64 extreme events (e.g., droughts, market fluctuations). In particular, the class of “vine” copulas
65 was developed to describe the dependence structure among multiple variables by building a
66 model based on a variety of joint bivariate distribution functions (Czado and Nagler, 2022).
67 Copula models have been applied to a wide range of topics, such as meteorological events and
68 economics, for decades. For instance, copula models were used to describe how the dependence
69 between rainfall and temperature influences agricultural production, to help study effects of
70 climate change on crop yields (Cong and Brady, 2012). Likewise, copula models were recently
71 used to describe the dependence structure among distributions of soil parameters that traditional
72 multivariate normal distributions were unable to represent (Lü et al., 2020).

73 The aim of the present study was to model the pairwise dependence between a regulating SES
74 (groundwater recharge or soil carbon (C) sequestration) and a provisioning SES (plant biomass
75 provision). We then investigated the potential to increase the level of each of the regulating
76 SESs. The pairwise dependence for the two regulating SESs was also analyzed as a function of
77 a weather condition (i.e., effective rainfall (ER), equal to rainfall minus potential
78 evapotranspiration) or a soil property (i.e., soil organic C (SOC) content), respectively. Rather
79 than being measured in the field, the levels of the three SESs were predicted over 31 years using
80 the STICS soil-crop model based on a dataset of 64 cultivated soils in northwestern France.

81 STICS is a one-dimensional mechanistic model that simulates crop development and soil
82 processes that connect water, nitrogen (N) and C dynamics in the soil-plant-atmosphere
83 continuum (Brisson et al., 2009). STICS has a daily time step and can simulate multiple
84 consecutive years. Previous studies (e.g., Brisson et al. (2002); Schnebelen et al. (2004);
85 Constantin et al. (2012); Constantin et al. (2015)) have evaluated the accuracy with which
86 STICS predicted some of these outputs (including atypical values) for a variety of field and
87 cover crops. A review of studies that used STICS to simulate 15 field crops at a total of 76 sites
88 in France (Coucheney et al., 2015) considered that its accuracy was “very good” for whole-
89 profile soil water content (mean relative root mean square error (rRMSE) = 10%) and “good”
90 for plant fruit biomass at harvest (i.e., yield) (mean rRMSE = 33%). They concluded that STICS
91 had sufficient accuracy and robustness for large-scale use under the soil and climate conditions
92 in France. Similarly, predictions of changes in SOC content for seven long-term field
93 experiments in Europe using the AMG model, on which STICS bases its simulation of SOC
94 dynamics, were considered sufficiently accurate (rRMSE = 28%) (Levavasseur et al., 2020).
95 We expected that by simulating the same crop rotation and weather for all 64 soil profiles,

96 STICS would be able to predict relative differences among the profiles with sufficient accuracy
97 for each of the three SESs studied.

98

99 **2. Materials and Methods**

100 **2.1. Data**

101 **Soil dataset.** The soil dataset contained data from 64 sites of cultivated land sampled in the
102 department of Ille-et-Vilaine, in Brittany, northwestern France. Soil samples were collected at
103 multiple depth intervals according to *GlobalSoilMap* specifications (Arrouays et al., 2014), and
104 physico-chemical analyses of the samples were performed to measure properties such as SOC
105 content (dry combustion) and pH (1:5 soil-to-water ratio, NF ISO 10390). See descriptive
106 statistics of ER and SOC content (Table 1) and their scatterplots (Figure A1 in Supplementary
107 material) of the soil profiles of the 64 sites from 1988-2018. See Ellili-Bargaoui et al. (2021)
108 for more details.

109 **Weather data.** The Ille-et-Vilaine department has a temperate oceanic climate. Weather
110 conditions were assumed to be the same for all 64 sites in the study area, an assumption
111 supported by climate data from Météo France for 1981-2010 (Bretagne Environnement, 2020),
112 which showed that, for nearly all sites, annual rainfall varied by no more than 200 mm (i.e.,
113 700-900 mm), and mean annual temperature varied by no more than 1°C (i.e., 11-12°C).
114 Weather data came from the weather station of Rennes-St Jacques, located near the center of
115 the study area. Data such as daily rainfall (mm), mean air temperature (°C) and solar radiation
116 (kW/m²) were collected.

117 **Crop-management data.** Crop-management data reflected the main conventional crop rotation
118 used by farmers in the study area: a 2-year rotation of grain maize (10 Apr-30 Oct), winter
119 wheat (1 Nov-31 Jul) and a catch crop of white mustard (5 Sep-3 Mar). For grain maize and
120 wheat, the soil was plowed to a depth of ca. 25 cm during the first five days of the period, and
121 organic and inorganic fertilization was adjusted based on crop requirements, soil supplies and
122 requirements of the European Union Nitrates Directive (EC, 1991) as specified in the French
123 National Action Plan (OJFR, 2011).

124 **2.2. Simulation modeling with STICS**

125 Input parameters for STICS (version 9.0) came from the soil, weather and crop-management
126 data. During the simulated study period, all soil parameters remained fixed except SOC, N and
127 water contents, since they were influenced by crop development and weather conditions. STICS
128 was used to simulate sequentially the three crops in the rotations at the 64 sites for 31
129 agricultural years (1988-2018) with the aid of a Java package developed to automate
130 simulations. STICS provided more than 200 outputs, of which four daily outputs were selected
131 to calculate indicators of the SESs: crop transpiration, crop yield, water drainage and SOC in
132 humified organic matter. See Ellili-Bargaoui et al. (2021) for more details.

133 **2.3. Soil ecosystem services**

134 STICS outputs were used to estimate the level of the indicator of each SES (i.e., climate
135 regulation, groundwater recharge, and plant biomass provision) for each calendar year
136 simulated.

137 a. *Climate regulation* was estimated through C sequestration (CS), which represented the
138 amount of organic C that the soil stored or released (positive or negative value,
139 respectively). CS was quantified as the annual change in the stock of SOC in the topsoil
140 (the top 30 cm):

$$CS_j = CStock_j - CStock_{j-1} \quad \text{Equation 1}$$

141 where $CStock_j$ is the SOC stock in the topsoil in year j .

142 b. *Groundwater recharge* (GW) is the amount of water that percolates into the
143 groundwater. GW was quantified as the annual sum of the water drained daily through the
144 soil, which was assumed to reach the groundwater:

$$GW_j = \sum_{i=1}^{365} D_i \quad \text{Equation 2}$$

145 where i is the day of year (1-365), and D_i is the amount of water drained daily from the soil
146 (mm).

147 c. *Plant biomass provision* (YE) represents the soil's ability to produce plant biomass by
148 photosynthesis. YE equals the annual yields of cash crops expressed as a unit of energy (GJ
149 $\text{ha}^{-1} \text{year}^{-1}$):

$$YE_j = \frac{1}{n} \sum_{i=1}^n B_i \times k$$

Equation 3

150 where B_i is the dry matter ($\text{t ha}^{-1} \text{ year}^{-1}$) of harvested biomass, and k is the energy content of
151 biomass: 14.905 and 13.984 GJt^{-1} fresh matter for maize and wheat grain, respectively (FAO,
152 2001).

153 After simulation modeling, the SES dataset contained 5952 variables for the SESs studied (3
154 SESs \times 64 sites \times 31 years of simulation). Sources of variation when predicting SESs
155 corresponded to differences in weather conditions among years and to soil properties among
156 sites, as the succession of field interventions was simulated on the same dates each year over
157 the study period and for each soil profile. See descriptive statistics (Table 1) and scatterplots
158 (Figure A1) for the three SESs of the soil profiles of the 64 sites from 1988-2018.

159 **Table 1.** Descriptive statistics of effective rainfall (ER, rainfall minus potential
 160 evapotranspiration) and soil organic carbon (SOC) content and predicted soil ecosystem
 161 services (groundwater recharge and plant biomass provision) of the 64 cultivated soil profiles
 162 (Ellili-Bargaoui et al., 2021), and for samples of the lowest (10th percentile) or highest (90th
 163 percentile) of ER and SOC contents

Variable	Data	Minimum	Mean	Maximum	Standard deviation
Effective rainfall (mm yr ⁻¹)	Entire dataset	-440	-256	-51	122
	Lowest ER	-440	-428	-417	8
	Highest ER	-97	-73	-51	15
SOC content (g/kg)	Entire dataset	10	20	45	7
	Lowest SOC contents	10	11	12	0.6
	Highest SOC contents	26	34	45	8
Plant biomass provision (GJ ha ⁻¹ yr ⁻¹)	Entire dataset	2	82	187	30
	Lowest ER	24	55	99	14
	Highest ER	39	87	179	28
	Lowest SOC contents	2	94	187	32
	Highest SOC contents	24	62	137	23
Groundwater recharge (mm yr ⁻¹)	Entire dataset	0	182	451	99
	Lowest ER	0	99	227	63
	Highest ER	76	288	441	74
	Lowest SOC contents	0	151	384	92
	Highest SOC contents	3	191	451	103
Climate regulation (kg C ha ⁻¹ yr ⁻¹)	Entire dataset	-1828	126	1279	344
	Lowest ER	-1454	199	1085	8
	Highest ER	-896	180	1279	15
	Lowest SOC contents	-602	189	950	263
	Highest SOC contents	-1828	-275	448	442

164

165 **2.4. C-Vine Copula method**

166 This section provides an overview of the copula method for modeling the dependence structure
 167 among three variables. Unlike other statistical methods, the vine copula method provides many
 168 functions to represent how three variables vary simultaneously by decomposing them into
 169 multiple bivariate functions. For more details, many articles review copulas comprehensively
 170 (Aas et al., 2009; Nagler, 2014; Nadarajah et al., 2018).

171 **2.4.1. Vine copula decomposition**

172 We consider continuous random variables X, Y and Z . Their marginal cumulative distribution
 173 functions (cdf) are denoted by F_X, F_Y and F_Z (i.e., probability of values falling below a
 174 threshold such that $F_X(x) = \Pr(X \leq x)$ (range: $[0, 1]$)). Their probability density functions
 175 (pdf) are denoted by f_X, f_Y and f_Z (i.e., probability of falling within a particular range of values
 176 based on the shape of the distribution such that $\Pr(a \leq X \leq b) = \int_a^b f(x)dx$ (range: $[0, 1]$)).
 177 Then, the triplet of variables (X, Y, Z) has a joint cdf $F(x, y, z) = \Pr(X \leq x, Y \leq y, Z \leq z)$ and
 178 a joint pdf denoted by f , which describes the simultaneous distribution of X, Y and Z . The
 179 cdfs F_X, F_Y, F_Z are unknown functions to estimate from data with the corresponding densities.
 180 We also denote the copula cdf as C and the copula pdf as c .

181 Given real variables X, Y and Z , a copula is a function that associates the triplet
 182 $(F_X(x), F_Y(y), F_Z(z))$ in $[0, 1]^3$ of marginal cdfs to the joint cdf $F(x, y, z)$ in $[0, 1]$, as follows
 183 (Sklar, 1959):

$$C(F_X(x), F_Y(y), F_Z(z)) = F(x, y, z) \quad \text{Equation 4}$$

184 By connecting F to F_X, F_Y and F_Z , the copula function C describes the dependence among X, Y
 185 and Z , if it exists. ‘‘Canonical’’ vine (C-vine) copulas decompose the joint pdf f using various
 186 pair-copulas, such as

$$f(x, y, z) = f_X(x) \cdot f_Y(y) \cdot f_Z(z) \quad \text{Equation 5}$$

$$\cdot c_{xy}(F_X(x), F_Y(y)) \cdot c_{yz}(F_Y(y), F_Z(z)) \quad (\text{T}_1)$$

$$\cdot c_{xz|y}(F(x|y), F(z|y)), \quad (\text{T}_2)$$

187 where copulas c_{xy}, c_{yz} and $c_{xz|y}$ join X and Y, Y and $Z, (X, Y)$ and (Y, Z) , respectively, and
 188 $F(x|y)$ and $F(z|y)$ are conditional cdf (Aas et al., 2009). A C-vine copula represents the
 189 multivariate relationship f among X, Y and Z as a set of trees (T), which consist of nodes (i.e.,
 190 variables) related by edges (i.e., bivariate copulas). Multiple decompositions can be developed
 191 depending on how the nodes are organized. Pair-copula decompositions are flexible tools that
 192 are useful for modeling dependence among variables of a high-dimensional model.

193 **2.4.2. Bivariate parametric copulas**

194

195 **2.4.2.1. Examples**

196 A variety of parametric copula classes (e.g., Archimedean, elliptical, extreme-value) have been
197 developed to model various dependence structures between variables, such as positive or
198 negative dependence (Nadarajah et al., 2018). Here, among the various copulas tested to fit the
199 dependence structure between SESs, we illustrate bivariate parametric copulas used in the vine
200 copula decomposition: the Archimedean copula “Frank” (Nagler, 2014) and the Gaussian
201 copula. We first denote the copula by $C_\theta(u, v)$, with a parameter θ , $u = F_X(x)$ and $v = F_Y(y)$
202 in $[0, 1]$.

203 *Example 1.* The Frank copula C_θ^{Frank} , with its real parameter $\theta \neq 0$, is defined by

$$C_\theta^{Frank}(u, v) = \frac{1}{\ln\theta} \ln \left(1 + \frac{(\theta^u - 1)(\theta^{-v} - 1)}{\theta - 1} \right) \quad \text{Equation 6}$$

204 The value of parameter θ provides information about the structure of copula C_θ^{Frank} : $\theta > 1$
205 corresponds to negative dependence, θ around 1 corresponds to independence, and θ in $]0, 1[$
206 corresponds to positive dependence.

207 *Example 2.* The Gaussian copula, with its parameter θ in $[-1, 1]$, is defined by

$$C_\theta^{Gauss}(u, v) = \frac{1}{2\pi\sqrt{1-\theta^2}} \int_{-\infty}^{\phi^{-1}(u)} \int_{-\infty}^{\phi^{-1}(v)} e^{\left(\frac{x^2-2\theta xy+y^2}{2(1-\theta^2)}\right)} dx dy \quad \text{Equation 7}$$

208 where ϕ^{-1} is the inverse of the univariate standard normal cdf. The value of parameter θ in
209 $[-1, 0[$ corresponds to negative dependence, θ around 0 corresponds to independence, and θ in
210 $]0, 1]$ corresponds to positive dependence.

211 **2.4.2.2. Copula selection and goodness-of-fit tests**

212 We considered independent realizations $(x_i, y_i), i = 1, \dots, n$, from the pair (X, Y) of continuous
213 random variables with their marginal cdf F_X and F_Y . The selection procedure first consisted of
214 fitting several copulas from the main copula classes. For each copula C_θ considered, an estimate
215 $\hat{\theta}_n$ of parameter θ was first calculated using the method of maximum likelihood such that

$$\hat{\theta}_n = \operatorname{argmax}_{\theta} \sum_{i=1}^n \ln c_{\theta}(u_i, v_i) \quad \text{Equation 8}$$

216 where c_{θ} is the copula pdf and (u_i, v_i) are the pairs obtained by transforming realizations
 217 (x_i, y_i) to uniform margins on $[0, 1]^2$, which must be done before applying Equation 8. We then
 218 selected the copula that minimized the Akaike Information Criterion (AIC), given by

$$AIC_n = -2 \sum_{i=1}^n \ln(c_{\hat{\theta}_n}(u_i, v_i)) + 2p \quad \text{Equation 9}$$

219 where p is the dimension of parameter θ in the copula ($= 1$ for the Frank and Gaussian copulas).

220 Once one copula $C_{\hat{\theta}_n}$ with the estimate $\hat{\theta}_n$ of its parameter was selected, the scatterplot of pairs
 221 of observations was compared to $N = 1000$ random pairs simulated using the Monte Carlo
 222 (MC) method from copula $C_{\hat{\theta}_n}$. The MC method is well adapted for simulating relationships
 223 among multiple dependent variables (Li et al., 2013), such as copulas, thus extrapolating a
 224 relatively small sample to a much larger one and increasing the robustness of the results
 225 obtained. Simulating random pairs from selected copula required finding the estimates \hat{F}_X and
 226 \hat{F}_Y of cdf F_X and F_Y . All analyses were performed in R software (R Core Team, 2020) using
 227 packages that applied the copula method, such as *VineCopula* (Schepsmeier et al., 2015).

228 **2.5. Application of the copula method**

229 We used results of copula models to estimate the probability of exceeding a level of a regulating
 230 SES as a function of the provisioning SES, by considering given weather conditions or soil
 231 properties. The C-vine copula was first applied to model joint cdfs of the three SESs (GW, CS
 232 and YE) calculated from STICS predictions, and classes of bivariate copulas that joined pairs
 233 of SESs were identified. Using maximum likelihood estimation, we tested lognormal, gamma,
 234 Weibull and normal pdfs to fit marginal distributions of each SES based on the Kolmogorov-
 235 Smirnov test (Figures A2 and A3). Conditional probabilities of one regulating SES (GW or CS)
 236 exceeding a threshold given that the provisioning SES (YE) exceeded another threshold were
 237 calculated (Box 1, Supplementary material). To this end, thresholds of GW, CS and YE were
 238 defined according to the range in which each SES varied in the soil dataset. Thresholds y of YE
 239 were varied from 20-100 GJ ha⁻¹ yr⁻¹, with a step of 20 GJ ha⁻¹ yr⁻¹, to consider a range of YE
 240 when considering the lowest or highest ER among all sites. For reference, mean (2000-2020)
 241 yields of maize grain and soft winter wheat in the study area (8.52 and 7.31 t ha⁻¹ of fresh matter,

242 respectively) (Agreste, 2021) would provide a mean of $114.6 \text{ GJ ha}^{-1} \text{ yr}^{-1}$ over the 2-year
243 rotation. Thresholds g of GW were varied from 0-250 and 75-500 mm yr^{-1} , with a step of 25
244 mm yr^{-1} , when considering the lowest or highest ER among all sites, respectively. Thresholds
245 c of CS were varied from -600 to 1000 $\text{kg C ha}^{-1} \text{ yr}^{-1}$ when considering the lowest or highest
246 ER among all sites, and from -1200 to 600 $\text{kg C ha}^{-1} \text{ yr}^{-1}$ or -600 to 600 $\text{kg C ha}^{-1} \text{ yr}^{-1}$ when
247 considering the lowest or highest SOC contents among all sites, respectively (with a step of 100
248 $\text{kg C ha}^{-1} \text{ yr}^{-1}$ for both). We focused in particular on the CS threshold of 0, at which soils began
249 to sequester SOC instead of releasing it.

250 Dependence among the three SESs was modeled for the lowest (10th percentile) or highest (90th
251 percentile) of ER and for the lowest (10th percentile) or highest (90th percentile) SOC contents
252 among all sites (Table 1). The percentiles of ER contained 199 values each (i.e. 10% of the total
253 sample of 64 sites \times 31 years = 1984). In comparison, the percentiles of SOC content contained
254 217 values each because the observed SOC content, used as the initial SOC content in the
255 simulations, remained the same for each site for all 31 simulated years (i.e., the total sample
256 contained 64 sets of 31 identical SOC contents; thus, no part of any set could be excluded from
257 a percentile). Conditional probabilities were calculated from the $N = 1000$ points simulated
258 from the selected copula using the MC method for the pairs of SESs considered (Equation A2).

259

260 **3. Results**

261 **3.1. 10th and 90th percentiles**

262 Because the same weather data were used for all 64 sites, estimates of ER for a given year
263 varied little among sites, due to small differences in predicted potential evapotranspiration
264 caused by differences in soil depth and predicted water-holding capacity. Thus, the 10th
265 percentile of ER contained all 64 sites for the three driest years (i.e., 1989, 2005 and 2010) and
266 7 sites for the next driest year (i.e. 1990). The 90th percentile of ER contained all 64 sites for
267 the three wettest years (i.e., 1998, 1999 and 2000) and 7 sites for the next wettest year (i.e.,
268 2014). In comparison, because SOC content is site-specific, and the observed SOC content was
269 used only to initialize the simulations, the 10th and 90th percentiles of SOC content each
270 contained 7 sites for all 31 simulated years. The soils with the lowest SOC contents among all
271 sites developed from recent marine alluvium (sandy to silty-loam deposits) or Aeolian loam.
272 Both parent materials provide conditions for low SOC content: nearly neutral or high pH and
273 great depth, fine texture and absence of coarse fragments, which allows for relatively intense

274 and deep plowing that may have led to relatively low SOC content. In contrast, most of the soils
275 with the highest SOC contents among all sites were in the Dol-de-Bretagne marsh and
276 developed from alternating continental and marine alluvia, which have naturally high clay and
277 SOC contents.

278 **3.2. Dependence among GW, CS and YE as a function of effective rainfall**

279 When fitting a C-vine copula to data of the three SESs studied, Frank copulas best fit the
280 dependence structure between GW and YE and CS and YE (Figure 1), but with different
281 estimated values $\hat{\theta}_n$ of the copula parameter for the lowest and highest ER. GW was negatively
282 correlated with YE (Kendall correlation coefficients $\hat{\tau}_n < 0$) when considering the lowest or
283 highest ER among all sites (Table 2) and the entire dataset (Figure A1). Conversely, CS was
284 positively ($\hat{\tau}_n > 0$) and negatively ($\hat{\tau}_n < 0$) correlated with YE when considering the lowest or
285 highest ER, respectively, compared to a positive correlation when considering the entire dataset.
286 Differing orientations and shapes of contour plots illustrated the sign (positive or negative) and
287 strength (high or low), respectively, of correlations between CS and YE as a function of ER
288 (Figure 1).

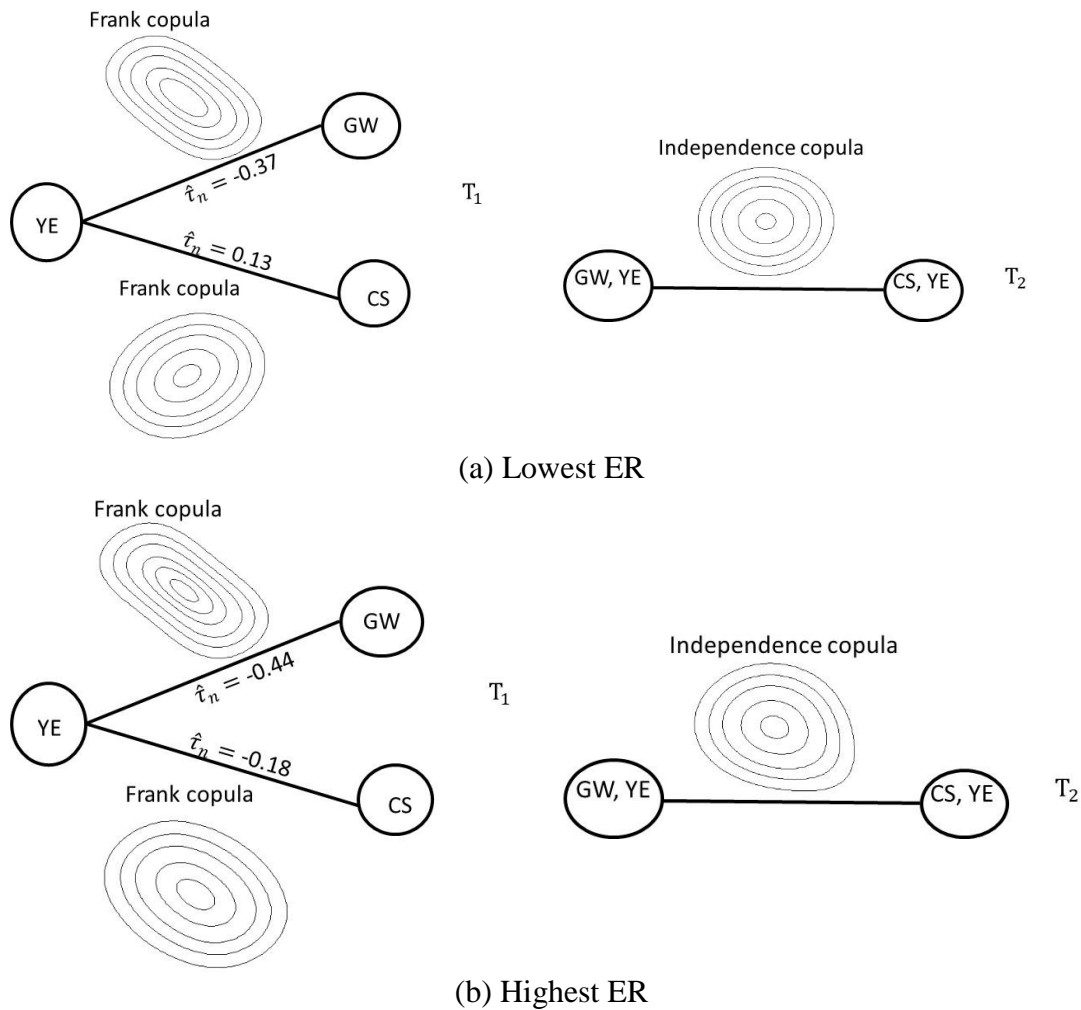
289 No tail dependences were identified between each pair of SESs due to the lack of soil profiles
290 that had simultaneously low or high values of GW or CS and YE (Figure 1). Conversely,
291 copulas modeled soil profiles in which GW or CS was lower when YE was higher, and vice-
292 versa. C-vine copulas also modeled an independence structure (i.e., a more circular contour)
293 between the pairs (GW, YE) and (CS, YE) when considering either the lowest or highest ER
294 (Figure 1). Thus, YE influenced GW and CS separately, but did not influence the dependence
295 between GW and CS.

296

297 **Table 2.** Kendall correlation coefficients $\hat{\tau}_n$ between pairs of soil ecosystem services (SESs)
 298 (climate regulation (CS), groundwater recharge (GW), and plant biomass provision (YE))
 299 and/or weather condition (effective rainfall (ER)) and soil property (organic carbon (SOC)
 300 content) of soil profiles associated with the lowest or highest ER and lowest or highest SOC
 301 contents among all sites. Bold values indicate significant correlations ($p < 0.05$).

SES	Lowest ER				Highest ER			
	CS	GW	YE	ER	CS	GW	YE	ER
CS	1.00	-0.03	0.13	-0.30	1.00	0.07	-0.19	-0.31
GW		1.00	-0.37	0.17		1.00	-0.44	0.05
YE			1.00	-0.06			1.00	0.19
SES	Lowest SOC content				Highest SOC content			
	CS	GW	YE	SOC	CS	GW	YE	SOC
CS	1.00	0.06	-0.10	0.20	1.00	0.25	-0.07	-0.49
GW		1.00	0.01	0.08		1.00	-0.05	-0.14
YE			1.00	-0.22			1.00	-0.08

302



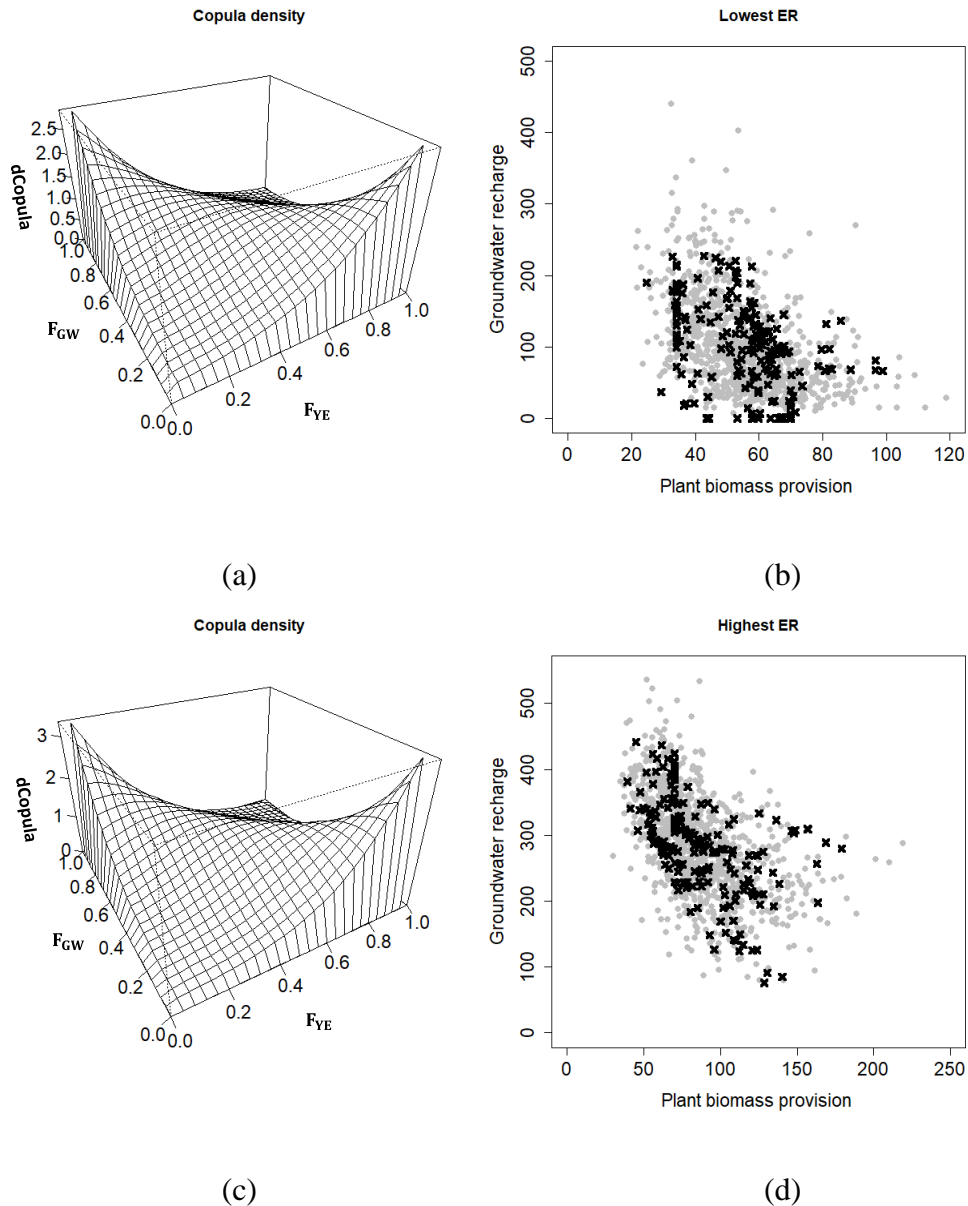
303 **Figure 1.** Trees (T_1 and T_2) of canonical vine copulas and contours of bivariate copulas selected
 304 to model dependence among groundwater recharge (GW), carbon sequestration (CS) and plant
 305 biomass provision (YE) calculated from predictions of the STICS model for sites with the (a)
 306 lowest effective rainfall (ER) or (b) highest ER. $\hat{\tau}_n$ is the Kendall correlation coefficient.

307

308 3.2.1. Groundwater recharge as a function of plant biomass provision

309 The dependence between GW and YE modeled by Frank copulas was investigated first. The
 310 two scatterplots of 199 points calculated from STICS predictions for sites with the lowest or
 311 highest ER and of 1000 points extrapolated from the fitted copulas overlapped almost
 312 completely, which indicated that these copulas represented the predictions satisfactorily (Figure
 313 2b, d).

314



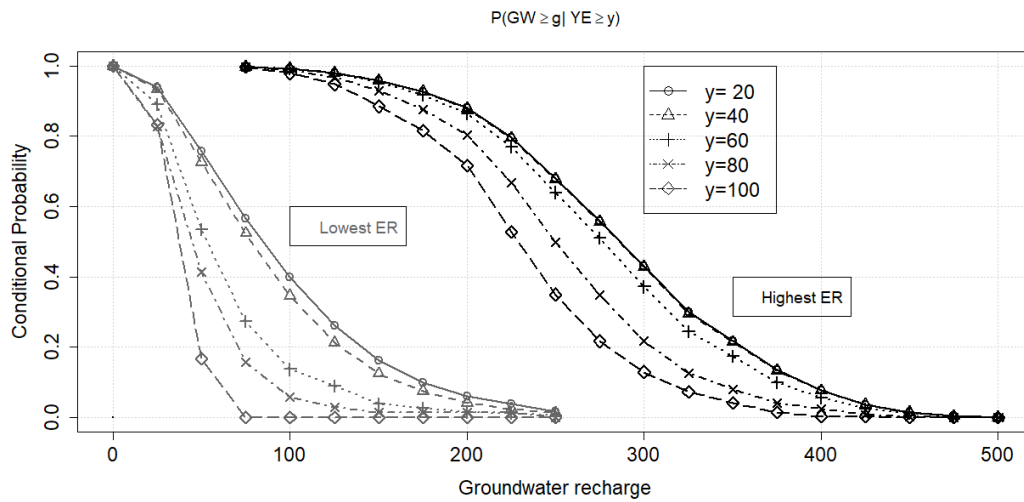
315 **Figure 2.** (a, c) Density functions of Frank copulas selected to model the dependence between
 316 univariate marginal cumulative distribution functions of groundwater recharge (GW, mm yr⁻¹)
 317 and plant biomass provision (YE, GJ ha⁻¹ yr⁻¹) calculated from predictions of the STICS model
 318 for sites with the (a, b) lowest effective rainfall (ER) or (c, d) highest ER extrapolated from a
 319 random sample (gray circles) of $N = 1000$ from the fitted copula. Black crosses represent 199
 320 points from STICS predictions.

321

322 Based on values extrapolated from the chosen copula, the probability of GW exceeding a given
 323 threshold logically decreased as this threshold increased from 0 to 250 mm yr⁻¹ for the lowest
 324 ER or from 75 to 500 mm yr⁻¹ for the highest ER, for a given threshold of YE (Figure 3, Tables
 325 A1 and A2). In addition, as the threshold of YE increased from 20 to 100 GJ ha⁻¹ yr⁻¹, the

326 probability of GW exceeding a given threshold decreased. For instance, as the threshold of YE
 327 increased from 20 to 60 GJ ha⁻¹ yr⁻¹, the probability of exceeding 100 mm yr⁻¹ for the lowest
 328 ER decreased from 0.42 to 0.16 and that of exceeding 300 mm yr⁻¹ during the highest ER
 329 decreased from 0.43 to 0.24.

330 When varying thresholds of YE, the probability of exceeding a given threshold of GW varied
 331 more when considering the lowest ER than the highest ER. The threshold of YE had little
 332 influence on the probability of GW exceeding a minimum or maximum threshold, but it had a
 333 strong influence on the probability of GW exceeding intermediate thresholds. The variability
 334 in probability was particularly high for certain thresholds of GW for the lowest or highest ER.
 335 For instance, the probability of exceeding GW thresholds of 50 or 100 mm yr⁻¹ for the lowest
 336 ER (the latter close to the mean for these ER (Table 1)) was 0.25-0.77 and 0-0.42, respectively,
 337 as the threshold of YE varied. Likewise, the probability of exceeding a GW threshold of 200 or
 338 300 mm yr⁻¹ during the highest ER (the latter close to the mean for these ER (Table 1)) was
 339 0.78-0.89 and 0.17-0.43, respectively, as the threshold of YE varied.



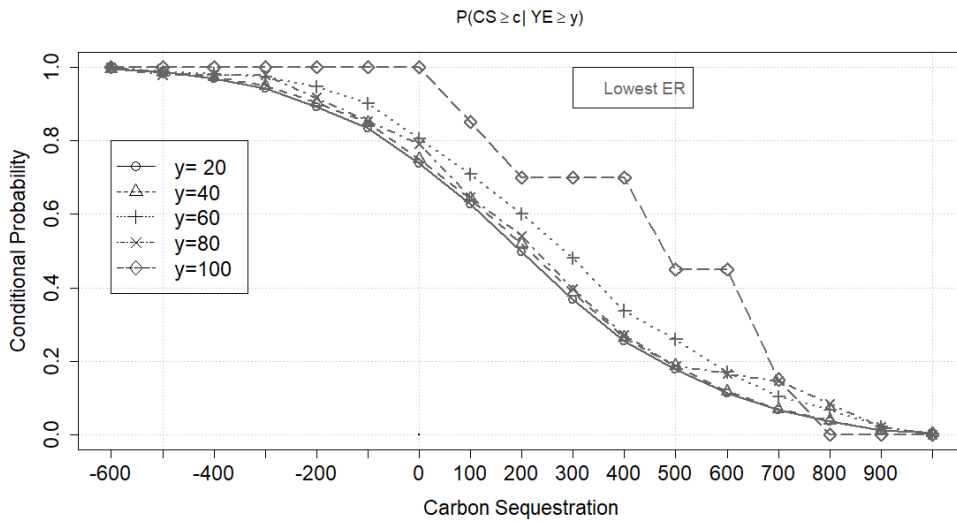
340
 341 **Figure 3.** Conditional probabilities of groundwater recharge (GW) (mm yr⁻¹) exceeding given
 342 thresholds (“g”) as a function of plant biomass provision (YE, GJ ha⁻¹ yr⁻¹) exceeding given
 343 thresholds (“y”) from extrapolation of STICS predictions for sites with the lowest effective
 344 rainfall (ER) (gray lines) or highest ER (black lines)

345

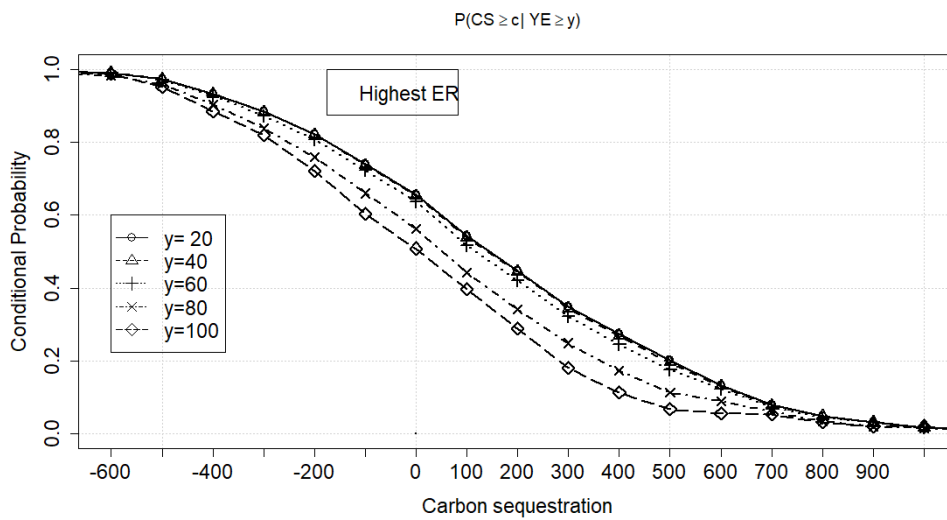
346 **3.2.2. Carbon sequestration as a function of plant biomass provision**

347 The conditional probability of CS exceeding a given threshold logically decreased as this
348 threshold increased from -600 to 1000 kg C ha⁻¹ yr⁻¹ for the lowest and highest ER, for a given
349 threshold of YE. As the threshold of YE increased from 20 to 100 GJ ha⁻¹ yr⁻¹, the probability
350 of CS exceeding a given threshold increased for the lowest ER (Figure 4a) but decreased for
351 the highest ER (Figure 4b). In addition, for the same thresholds of CS and YE, the probability
352 of CS exceeding a given threshold was generally higher for the lowest ER than for the highest
353 ER. For instance, the probability of exceeding the CS threshold of 0 was 0.70-1.00 for the
354 lowest ER and 0.50-0.70 for the highest ER.

355



(a)



(b)

356 **Figure 4.** Conditional probabilities of carbon sequestration (CS) ($\text{kg C ha}^{-1} \text{yr}^{-1}$) exceeding
 357 given thresholds (“c”) as a function of plant biomass provision (YE, $\text{GJ ha}^{-1} \text{yr}^{-1}$) exceeding
 358 given thresholds (“y”) from extrapolation of STICS predictions for sites with the (a) lowest
 359 effective rainfall (ER) or (b) highest ER

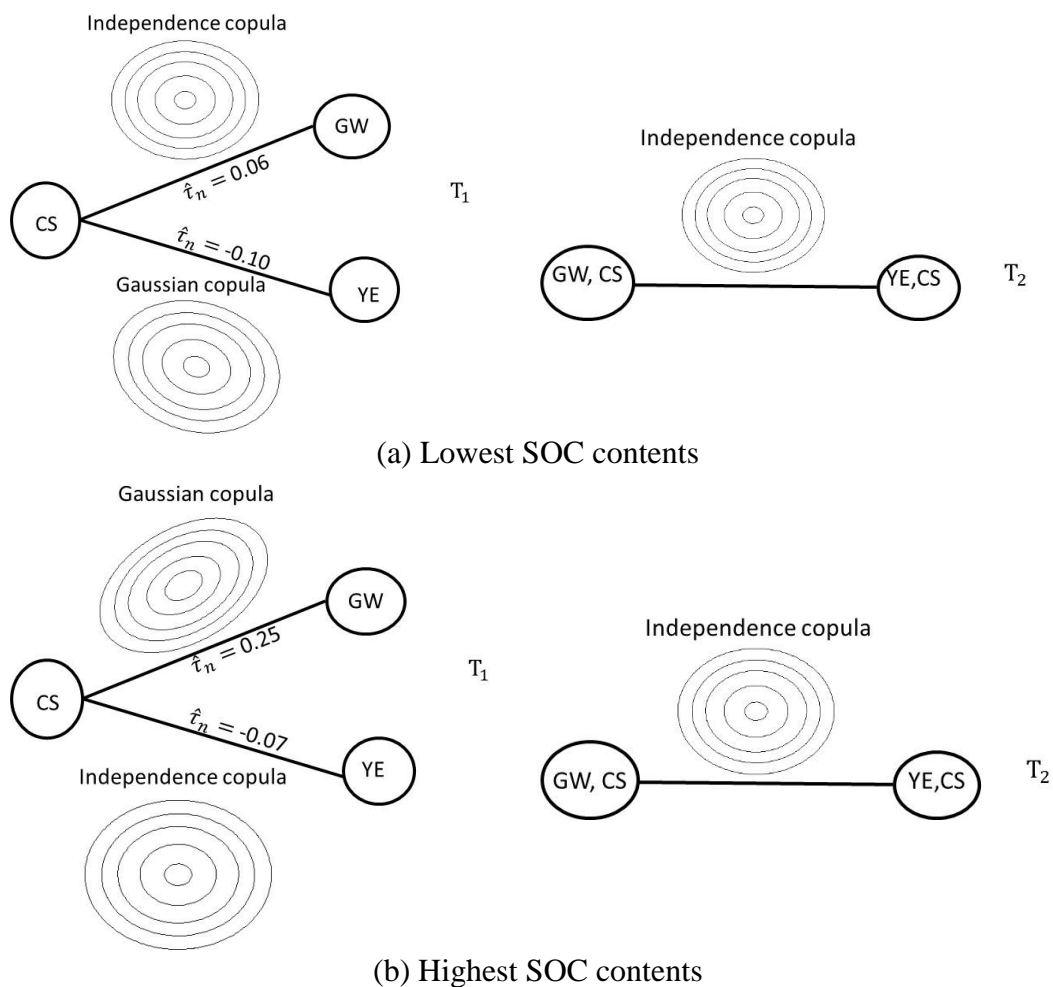
360

361 **3.3. Dependence among GW, CS and YE as a function of soil organic carbon** 362 **content**

363 When considering sites with the lowest SOC contents, the correlation between CS and YE was
 364 weak (-0.10) but significant (Table 1), and a Gaussian copula with $\hat{\theta}_n$ equal to -0.15 (i.e.,
 365 negative dependence) best fit the dependence between them (Figure 5). The correlation between

366 CS and GW was also weak (0.06) but not significant, and an independence copula fit the
 367 relationship between these two SES. Likewise, an independence copula fit the relationship
 368 between (GW, CS) and (YE, CS), which indicated that CS did not influence the relationship
 369 between GW and YE. When considering sites with the highest SOC contents, the correlation
 370 between CS and GW was significant (0.25), and a Gaussian copula with $\hat{\tau}_n$ equal to 0.39 fit the
 371 dependence between them. An independence copula fit the relationship between CS and YE,
 372 as well as that between (GW, CS) and (YE, CS) (Figure 5).

373

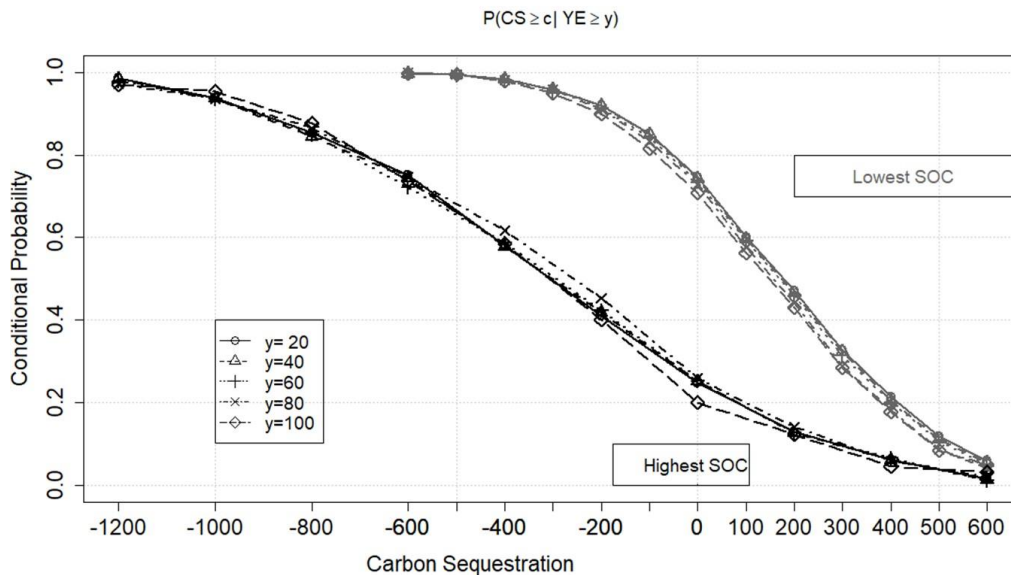


374 **Figure 5.** Trees (T_1 and T_2) of canonical vine copulas and contours of bivariate copulas selected
 375 to model the dependence among groundwater recharge (GW), carbon sequestration (CS) and
 376 plant biomass provision (YE) calculated from predictions of the STICS model for sites with the
 377 (a) lowest or (b) highest organic carbon (SOC) contents. $\hat{\tau}_n$ is the Kendall correlation
 378 coefficient.

379

380 **3.3.1. Carbon sequestration as a function of plant biomass provision**

381 The modeled dependence between CS and YE was then investigated. The two scatterplots of
 382 217 points calculated from STICS predictions for sites with the lowest or highest SOC contents
 383 and of 1000 points extrapolated from the chosen copulas overlapped almost completely (Figure
 384 A6). Like the dependence between GW and YE, the conditional probability of CS exceeding a
 385 given threshold logically decreased as this threshold increased from -600 to 600 kg C ha⁻¹ yr⁻¹
 386 for the lowest SOC contents or from -1200 to 600 kg C ha⁻¹ yr⁻¹ for the highest SOC contents,
 387 for a given threshold of YE (Figure 6, Tables A3 and A4). The conditional probability of CS
 388 exceeding a given threshold was influenced slightly or not all by the YE threshold, which
 389 reflected the weak correlation or independence between CS and YE for the lowest or highest
 390 SOC contents, respectively. For instance, the probabilities of exceeding CS thresholds of 0 and
 391 300 kg C ha⁻¹ yr⁻¹ for the lowest SOC contents were 0.71-0.75 and 0.28-0.33, respectively, as
 392 the threshold of YE varied. Likewise, the probabilities of exceeding the same two thresholds
 393 for the highest SOC contents were 0.21-0.27 and 0.07-0.09, respectively, as the threshold of
 394 YE varied. Thus, for the same thresholds of CS and YE, the conditional probability of CS
 395 exceeding a given threshold was higher for sites with the lowest SOC contents than for those
 396 with the highest SOC contents.



397
 398 **Figure 6.** Conditional probabilities of carbon sequestration (CS) exceeding given thresholds
 399 (“c”) (kg C ha⁻¹ yr⁻¹) as a function of plant biomass provision (YE, GJ ha⁻¹ yr⁻¹) exceeding given
 400 thresholds (“y”) from extrapolation of STICS predictions for sites with the (a) lowest (gray
 401 lines) or (b) highest (black lines) soil organic carbon (SOC) contents.

402

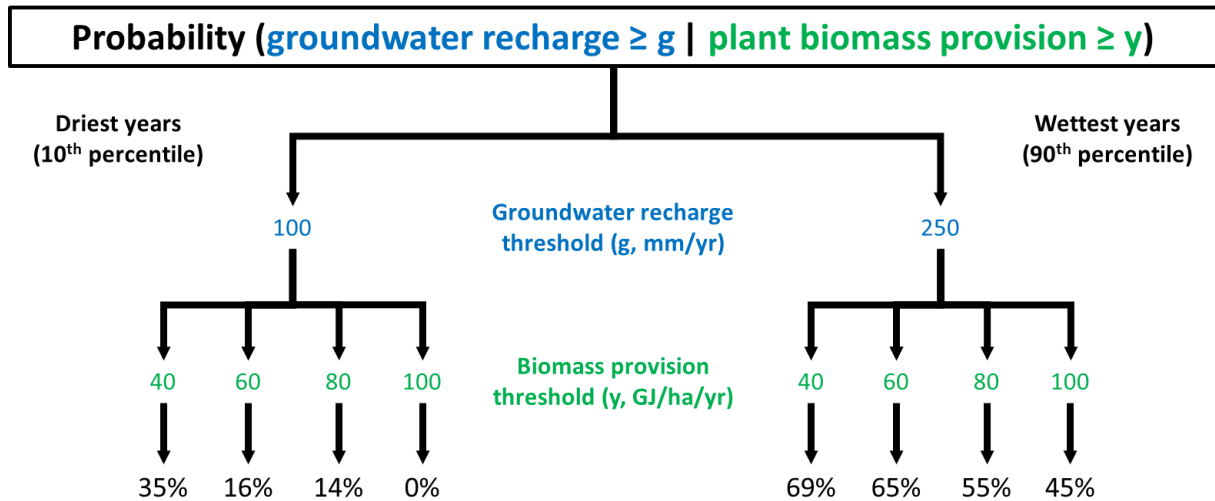
403 **4. Discussion**

404 **4.1. Probabilities of exceeding ecosystem service thresholds**

405 **4.1.1. Groundwater recharge as a function of plant biomass provision**

406 Results of copula modeling helped analyze the potential to achieve certain target levels of
407 regulating SESs (GW and CS) as a function of variations in a provisioning SES (YE), weather
408 conditions and a soil property (i.e., SOC content). The dependence that we observed between
409 YE and GW is well documented in the literature and can be explained by considering the soil
410 water balance (Radcliffe and Simunek, 2012). Higher crop yields increase the flow of water to
411 the atmosphere via transpiration, which decreases soil water content and the amount of water
412 likely to drain below the root system. Thus, most studies that analyzed the relationship between
413 YE and GW identified a significantly negative correlation between them (e.g., -0.71 in southern
414 France (Demestihis et al., 2018), -0.86 for the mean annual data considered in the present study
415 (Figure A5) (Ellili-Bargaoui et al., 2021)). Their magnitude, however, depends on annual
416 weather conditions, as shown by applying the copula method. The relationship between YE and
417 GW also depended on their ranges of variation, with lower conditional probabilities of
418 exceeding a given level of GW during the wettest years (i.e., less variation) than the driest years
419 (i.e., more variation) (Figure 3). Placing these conditional probabilities in a decision tree, GW
420 was more likely to exceed thresholds close to its mean value (i.e., 100 and 250 mm yr⁻¹ in the
421 wettest and driest years, respectively (Table 1)) during the wettest years than the driest years,
422 when considering YE exceeding thresholds of 40, 60 and 80 GJ ha⁻¹ yr⁻¹ (Figure 7). Thus, when
423 GW becomes an issue for the supply of drinking water or preservation of aquatic environments,
424 such as after a dry winter, certain management practices (e.g., decreasing fertilization) could be
425 used as a mechanism to limit YE in order to ensure sufficient levels of GW. The decision tree
426 based on copula models provides indications of the yield that could be targeted depending on
427 the level of GW desired, and the associated conditional probability.

428



429

430 **Figure 7.** A decision tree of conditional probabilities of groundwater recharge (GW) exceeding
 431 a given threshold ‘g’ (mm yr⁻¹) as a function of plant biomass provision (YE) exceeding a given
 432 threshold ‘y’ (GJ ha⁻¹ yr⁻¹) from extrapolation of STICS predictions for sites with the lowest or
 433 highest effective rainfall (i.e., driest or wettest years, respectively). For reference, mean yields
 434 of maize grain and wheat (8.52 and 7.31 t ha⁻¹, respectively) in the study area would provide a
 435 mean of 114.6 GJ ha⁻¹ yr⁻¹ over the 2-year rotation.

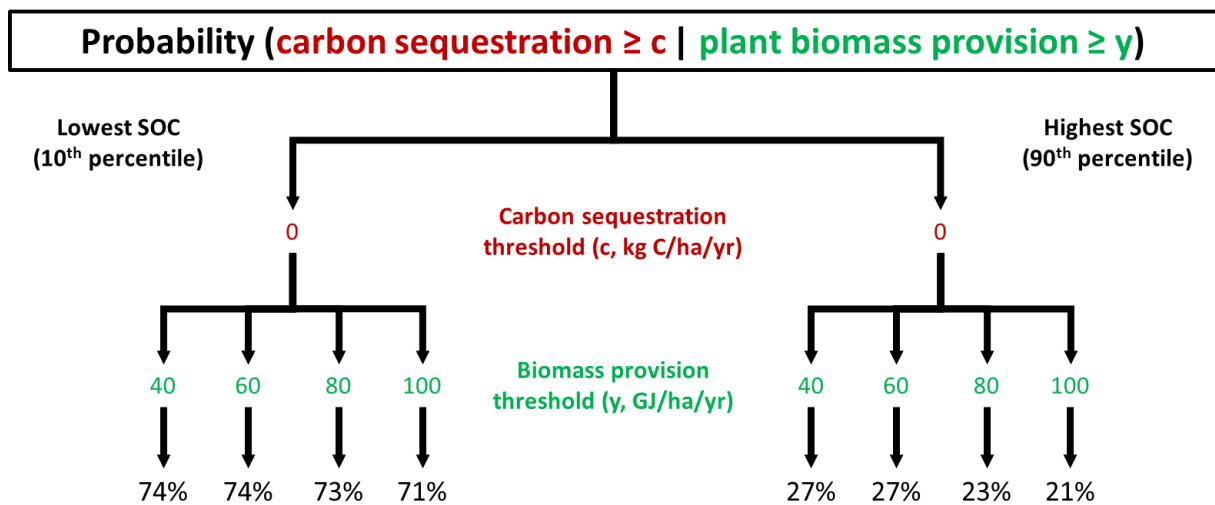
436

437 4.1.2. Carbon sequestration as a function of plant biomass provision

438 The dependence that we observed between yield YE and CS is more complex, since CS depends
 439 on the initial SOC and the type of crop considered, but also on management practices such as
 440 fertilization and the removal or incorporation crop residues (Lemke et al., 2010; Paustian,
 441 2014). The predictions used in our study assumed that all crop residues were removed. The
 442 threshold of YE influenced the probability of CS exceeding a given threshold when ER varied
 443 (i.e., driest vs. wettest years) (Figure 4). The CS may have been higher during the driest years
 444 in part because dry conditions decrease mineralization of organic matter by microorganisms
 445 (Thapa et al., 2021). Mineralization may also explain why, as YE increased, CS increased
 446 during the driest years but decreased during the wettest years. Temperature was not considered
 447 as a variable in the copula models, but it also influences mineralization in the field and the
 448 STICS model. Conversely, the threshold of YE had almost no influence on the probability of
 449 CS exceeding a given threshold when the SOC content in the topsoil varied (i.e., lowest vs.
 450 highest) (Figure 6). In this context, modifying YE may or may not be used as a mechanism to
 451 increase the probability of CS exceeding a given threshold (Figure A4).

452 YE was weakly negatively correlated with SOC (-0.22) (Figure A1). A soil's capacity to
 453 sequester C (i.e., positive CS) depends on its physicochemical properties, especially clay
 454 content (Churchman et al., 2020), and negative CS (i.e., loss of SOC) is considered a
 455 “disservice” (Olson et al., 2017). The probabilities that CS exceeded high thresholds (e.g., 300
 456 kg C ha⁻¹ yr⁻¹) were low, since soils tend towards saturation in SOC content. Consequently, soils
 457 were more likely to sequester C when considering the lowest SOC contents (0.71-0.75) than
 458 the highest SOC contents (0.21-0.27), regardless of the threshold of YE (Figure 8).

459



460

461 **Figure 8.** A decision tree of conditional probabilities of carbon sequestration (CS) exceeding a
 462 given threshold ‘c’ (kg C ha⁻¹ yr⁻¹) as a function of plant biomass provision (YE) exceeding a
 463 given threshold ‘y’ (GJ ha⁻¹ yr⁻¹) from extrapolation of STICS predictions for sites with the
 464 lowest or highest soil organic carbon (SOC) contents. For reference, mean yields of maize grain
 465 (7.66 t DM ha⁻¹) and wheat (6.48 t DM ha⁻¹) in the study area would provide a mean of 96.8 GJ
 466 ha⁻¹ yr⁻¹ over the 2-year rotation.

467

468 Trade-offs may need to be made between the threshold of a regulating SES (GW or CS) and
 469 the certainty that it can be achieved: more certainty that a lower threshold can be achieved or,
 470 conversely, less certainty that a higher threshold can be achieved. The influence of weather
 471 conditions and soil properties must also be considered when estimating probabilities that
 472 alternative options can achieve target levels of SES.

473 **4.2. The utility of vine copulas compared to those of other statistical methods**

474 Using the same dataset, Ellili-Bargaoui et al. (2021) developed a correlation network chart to
475 visualize relationships among the mean annual provision of YE, GW and CS and other SESs
476 for all 64 sites over the 31-year simulation (Figure A5). It illustrated a strong negative
477 correlation between YE and GW (Pearson correlation coefficient $r = -0.86$) and no significant
478 correlation between YE and CS ($r = 0.13$). When considering soil properties (e.g., SOC
479 content), the correlation network chart illustrated that SOC had a significantly negative
480 correlation with YE ($r = -0.45$) and CS ($r = -0.66$) but no significant correlation with GW
481 ($r = 0.16$). While Ellili-Bargaoui et al. (2021) used PCA to identify relations among groups of
482 SES, our use of vine copulas specifically enabled us to model degrees of dependence among
483 SESs, such as simple dependence between two SESs and more complex dependence among
484 three or more SESs. In addition, copula models can be used to extrapolate the original sample
485 to a larger one by combining it with the MC method, unlike the correlation network chart and
486 PCA. Copulas enable one to model dependence, if it exists, between the highest or lowest values
487 of pairs of variables (Embrechts et al., 2002). Variables that have non-normal distributions and
488 the potential to have threshold effects on other variables at unusually high or low values (i.e.,
489 tail dependence) are particularly suited for analysis by the copula method.

490 **4.3. Limitations and perspectives of the study**

491 The representativeness of this study's results required assuming that STICS accurately
492 predicted the SESs studied. Because biophysical indicators of SESs are often difficult to
493 measure, soil processes are usually simulated to estimate them based on weather data and soil
494 properties. The lack of tail dependence for either pair of SESs studied and the weak correlation
495 between YE and CS or GW limited the added value of the copula method. The study could be
496 extended to investigate the dependence structure among additional SESs studied by Ellili-
497 Bargaoui et al. (2021) (e.g., water-to-plant provision, water quality regulation, N-to-plant
498 provision) as a function of other soil properties (e.g., pH and clay contents of the topsoil,
499 maximum rooting depth), after first identifying significant and strong correlations among these
500 variables to avoid redundant information. Furthermore, the probability of two SESs
501 simultaneously exceeding respective thresholds, and not only one SES doing so, as in the
502 present study, can be investigated given another SES exceeding a given threshold. We did not
503 analyze these probabilities, since the relationships of the two regulating SES as a function of

504 the provisioning SES best fit an independence copula. In addition, dynamic dependence among
505 SESs could also be studied by applying copula modeling to time series of SESs.

506

507 **5. Conclusion**

508 We investigated how dependence between regulating and provisioning SESs can vary as a
509 function of their ranges of variation by applying vine copula models as a function of a weather
510 condition (i.e., ER) or soil property (i.e., SOC content in the topsoil). Although correlation
511 coefficients can quantify the strength and direction of linear relationships between SESs, they
512 are too simple to describe the complexity of this dependence. The vine copula models used
513 enabled us to formalize the dependence structure, if it existed, among the SESs studied and then
514 to extrapolate their joint variation from the original samples using the MC method. This
515 approach estimated the potential to achieve certain target levels of regulating SESs –
516 groundwater recharge and C sequestration – as a function of variations in a provisioning SES –
517 plant biomass provision. We found that contrasting weather conditions and a soil property could
518 influence the potential to improve one SES as a function of another SES. More complex
519 dependence among a bundle of SESs (i.e., four or more) could be studied using appropriate
520 multivariate copula models. Dynamic relationships among SESs could also be studied by
521 applying copula modeling to time series of SESs.

522

523 **Conflict of interest statement**

524 This manuscript has not been published and is not under consideration by another journal. The
525 authors have approved the manuscript and agree with submission to your esteemed journal.
526 There are no conflicts of interest to declare.

527

528 **Data availability statement**

529 Data sharing is not applicable to this article since no new data were created in this study.

530

531 **Authorship contribution statement**

532 Tristan Senga Kiessé: Conceptualization; formal analysis; methodology; visualization; writing
533 – original draft. Blandine Lemercier: Conceptualization; data curation; funding acquisition;
534 methodology; writing – review and editing. Michael S. Corson: Conceptualization;
535 methodology; visualization; writing – review and editing. Yosra Ellili-Bargaoui:

536 Conceptualization; data curation; funding acquisition; methodology; writing – review and
537 editing. Jihad Afassi: formal analysis; methodology. Christian Walter: Conceptualization; data
538 curation; funding acquisition; methodology; writing – review and editing.

539

540 **Acknowledgments**

541 We thank the INRAE division “Agronomy and environmental sciences for agroecosystems”
542 (AgroEcoSystem) for its financial support. This research was also developed in the framework
543 of the European Joint Program for SOIL "Towards climate smart sustainable management of
544 agricultural soils" (EJP SOIL), funded by the European Union Horizon 2020 research and
545 innovation programme (Grant Agreement N [862695](#)).

546 **References**

- 547 Aas, K., Czado, C., Frigessi, A., Bakken, H., 2009. Pair-copula constructions of multiple
548 dependence. *Insurance: Mathematics and economics* 44, 182-198.
- 549 Agreste, 2021. *Statistique agricole annuelle. Cultures développées (hors fourrage, prairies,*
550 *fruits, fleurs et vigne)*. Ministère de l'agriculture, [https://agreste.agriculture.gouv.fr/agreste-](https://agreste.agriculture.gouv.fr/agreste-web/disaron/SAANR_DEVELOPPE_2/detail/)
551 [web/disaron/SAANR_DEVELOPPE_2/detail/](https://agreste.agriculture.gouv.fr/agreste-web/disaron/SAANR_DEVELOPPE_2/detail/).
- 552 Arrouays, D., Grundy, M.G., Hartemink, A.E., Hempel, J.W., Heuvelink, G.B., Hong, S.Y., et
553 al., 2014. GlobalSoilMap: Toward a fine-resolution global grid of soil properties. *Advances in*
554 *agronomy* 125, 93-134.
- 555 Bretagne Environnement, 2020. [https://bretagne-environnement.fr/donnees-bilan-climatique-](https://bretagne-environnement.fr/donnees-bilan-climatique-bretagne)
556 [bretagne](https://bretagne-environnement.fr/donnees-bilan-climatique-bretagne).
- 557 Brisson, N., Launay, M., Mary, B., Beaudoin, N., 2009. Conceptual basis, formalisations and
558 parameterization of the STICS crop model. Editions Quae.
- 559 Brisson, N., Ruget, F., Gate, P., Lorgeou, J., Nicoullaud, B., Tayot, X., et al., 2002. STICS: a
560 generic model for simulating crops and their water and nitrogen balances. II. Model validation
561 for wheat and maize. *Agronomie* 22, 69-92.
- 562 Calzolari, C., Ungaro, F., Filippi, N., Guermandi, M., Malucelli, F., Marchi, N., et al., 2016. A
563 methodological framework to assess the multiple contributions of soils to ecosystem services
564 delivery at regional scale. *Geoderma* 261, 190-203.
- 565 Ching, J., Phoon, K.-K., Chen, C.-H., 2014. Modeling piezocone cone penetration (CPTU)
566 parameters of clays as a multivariate normal distribution. *Canadian Geotechnical Journal* 51,
567 77-91.
- 568 Churchman, G.J., Singh, M., Schapel, A., Sarkar, B., Bolan, N., 2020. Clay minerals as the key
569 to the sequestration of carbon in soils. *Clays and Clay Minerals* 68, 135-143.
- 570 Coles, S., Heffernan, J., Tawn, J., 1999. Dependence measures for extreme value analyses.
571 *Extremes* 24, 339-365.
- 572 Cong, R., Brady, M., 2012. The interdependence between rainfall and temperature: copula
573 analyses. *Sci World J* 2012: 1–11.
- 574 Constantin, J., Beaudoin, N., Launay, M., Duval, J., Mary, B., 2012. Long-term nitrogen
575 dynamics in various catch crop scenarios: test and simulations with STICS model in a temperate
576 climate. *Agriculture, Ecosystems & Environment* 147, 36-46.
- 577 Constantin, J., Le Bas, C., Justes, E., 2015. Large-scale assessment of optimal emergence and
578 destruction dates for cover crops to reduce nitrate leaching in temperate conditions using the
579 STICS soil–crop model. *European Journal of Agronomy* 69, 75-87.

580 Coucheney, E., Buis, S., Launay, M., Constantin, J., Mary, B., de Cortázar-Atauri, I.G., et al.,
581 2015. Accuracy, robustness and behavior of the STICS soil–crop model for plant, water and
582 nitrogen outputs: evaluation over a wide range of agro-environmental conditions in France.
583 *Environmental Modelling & Software* 64, 177-190.

584 Czado, C., Nagler, T., 2022. Vine copula based modeling. *Annual Review of Statistics and Its*
585 *Application* 9, 453-477.

586 Demestihias, C., Plénet, D., Génard, M., de Cortazar-Atauri, I.G., Launay, M., Ripoche, D., et
587 al., 2018. Analyzing ecosystem services in apple orchards using the STICS model. *European*
588 *Journal of Agronomy* 94, 108-119.

589 Dominati, E., Mackay, A., Lynch, B., Heath, N., Millner, I., 2014. An ecosystem services
590 approach to the quantification of shallow mass movement erosion and the value of soil
591 conservation practices. *Ecosystem Services* 9, 204-215.

592 EC, 1991. Council Directive 91/676/EEC of 12 December 1991 concerning the protection of
593 waters against pollution caused by nitrates from agricultural sources. *Official Journal of the*
594 *European Union* 375, 1-8.

595 Egoh, B., Reyers, B., Rouget, M., Richardson, D.M., Le Maitre, D.C., van Jaarsveld, A.S.,
596 2008. Mapping ecosystem services for planning and management. *Agriculture, Ecosystems &*
597 *Environment* 127, 135-140.

598 Ellili-Bargaoui, Y., Walter, C., Lemercier, B., Michot, D., 2021. Assessment of six soil
599 ecosystem services by coupling simulation modelling and field measurement of soil properties.
600 *Ecological Indicators* 121, 107211.

601 Embrechts, P., Mcneil, A., Straumann, D., 2002. Correlation and dependence in risk
602 management: properties and pitfalls. *Risk Management: Value at Risk and Beyond*. Cambridge
603 University Press, pp. 176--223.

604 FAO, 2001. Food balance sheets. A handbook. Food and Agricultural Organization of the
605 United Nations, Rome, p. 95.

606 Grace, J.B., 2006. Structural equation modeling and natural systems. Cambridge University
607 Press.

608 Joe, H., 1993. Parametric families of multivariate distributions with given margins. *Journal of*
609 *Multivariate Analysis* 46, 262-282.

610 Joly, F., Benoit, M., Martin, R., Dumont, B., 2021. Biological operability, a new concept based
611 on ergonomics to assess the pertinence of ecosystem services optimization practices. *Ecosystem*
612 *Services* 50, 101320.

613 Lee, H., Lautenbach, S., 2016. A quantitative review of relationships between ecosystem
614 services. *Ecological Indicators* 66, 340-351.

615 Lemke, R., VandenBygaart, A., Campbell, C., Lafond, G., Grant, B., 2010. Crop residue
616 removal and fertilizer N: effects on soil organic carbon in a long-term crop rotation experiment
617 on a Udic Boroll. *Agriculture, Ecosystems & Environment* 135, 42-51.

618 Leong, R.A.T., Fung, T.K., Sachidhanandam, U., Drillet, Z., Edwards, P.J., Richards, D.R.,
619 2020. Use of structural equation modeling to explore influences on perceptions of ecosystem
620 services and disservices attributed to birds in Singapore. *Ecosystem Services* 46, 101211.

621 Levavasseur, F., Mary, B., Christensen, B.T., Duparque, A., Ferchaud, F., Kätterer, T., et al.,
622 2020. The simple AMG model accurately simulates organic carbon storage in soils after
623 repeated application of exogenous organic matter. *Nutrient Cycling in Agroecosystems* 117,
624 215-229.

625 Li, D.-Q., Jiang, S.-H., Wu, S.-B., Zhou, C.-B., Zhang, L.-M., 2013. Modeling multivariate
626 distributions using Monte Carlo simulation for structural reliability analysis with complex
627 performance function. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal*
628 *of Risk and Reliability* 227, 109-118.

629 Lü, T.-J., Tang, X.-S., Li, D.-Q., Qi, X.-H., 2020. Modeling multivariate distribution of multiple
630 soil parameters using vine copula model. *Computers and Geotechnics* 118, 103340.

631 Mouchet, M.A., Lamarque, P., Martín-López, B., Crouzat, E., Gos, P., Byczek, C., Lavorel, S.,
632 2014. An interdisciplinary methodological guide for quantifying associations between
633 ecosystem services. *Global Environmental Change* 28, 298-308.

634 Nadarajah, S., Afuecheta, E., Chan, S., 2018. A compendium of copulas. *Statistica* 77, 279–
635 328.

636 Nagler, T., 2014. Kernel methods for vine copula estimation. Department of Mathematics.
637 Technische Universität München.

638 Nelson, E., Mendoza, G., Regetz, J., Polasky, S., Tallis, H., Cameron, D., et al., 2009. Modeling
639 multiple ecosystem services, biodiversity conservation, commodity production, and tradeoffs
640 at landscape scales. *Frontiers in Ecology and the Environment* 7, 4-11.

641 OJFR, 2011. Arrêté du 19 décembre 2011 relatif au programme d'actions national à mettre en
642 œuvre dans les zones vulnérables afin de réduire la pollution des eaux par les nitrates d'origine
643 agricole. <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000025001662>.

644 Olson, K.R., Al-Kaisi, M., Lal, R., Morton, L.W., 2017. Soil ecosystem services and intensified
645 cropping systems. *Journal of Soil and Water Conservation* 72, 64A-69A.

646 Paustian, K., 2014. Carbon Sequestration in Soil and Vegetation and Greenhouse Gases
647 Emissions Reduction. In: Freedman, B. (Ed.), *Global Environmental Change*. Springer
648 Netherlands, Dordrecht, pp. 399-406.

649 R Core Team, 2020. R: A language and environment for statistical computing. R Foundation
650 for Statistical Computing, Vienna, Austria.

651 Radcliffe, D., Simunek, J., 2012. Water flow in soils. In: Huang, P., Li, Y., Sumner, M. (Eds.),
652 *Handbook of soil sciences: properties and processes*. CRC Press Taylor and Francis Group, pp.
653 5.1-5.34.

654 Schepsmeier, U., Stoeber, J., Brechmann, E., et al., 2015. *VineCopula: Statistical Inference of*
655 *Vine Copulas*. R package version 1.6-1.

656 Schnebelen, N., Nicoullaud, B., Bourennane, H., Couturier, A., Verbeque, B., Revalier, C., et
657 al., 2004. The STICS model to predict nitrate leaching following agricultural practices.
658 *Agronomie* 24, 423-435.

659 Sklar, A., 1959. Fonctions de répartition à n dimensions et leurs marges. *Publications de*
660 *l'Institut de Statistique de l'Université de Paris* 8, 229-231.

661 Thapa, R., Tully, K.L., Cabrera, M.L., Dann, C., Schomberg, H.H., Timlin, D., et al., 2021.
662 Effects of moisture and temperature on C and N mineralization from surface-applied cover crop
663 residues. *Biology and Fertility of Soils* 57, 485-498.

664 Wu, J., Jin, X., Feng, Z., Chen, T., Wang, C., Feng, D., Lv, J., 2021. Relationship of Ecosystem
665 Services in the Beijing–Tianjin–Hebei Region Based on the Production Possibility Frontier.
666 *Land* 10, 881.

667

668