ELSEVIER

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv





Beyond a deterministic representation of the temperature dependence of soil respiration

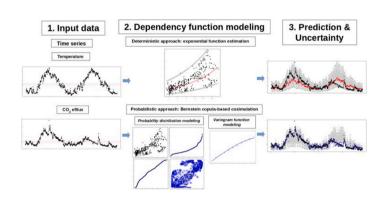
Van Huong Le¹, Rodrigo Vargas^{*,1}

Department of Plant and Soil Sciences, University of Delaware, Newark, DE, United States of America

HIGHLIGHTS

- We introduce the Bernstein copulabased cosimulation for modeling soil CO₂ efflux.
- The approach considers the joint probability distribution and temporal dependence of soil CO₂ efflux.
- The method improves accuracy and precision for prediction purposes.
- Probabilistic approaches hold promise for accurately representing dependency relationships.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Jay Gan

Keywords:
Soil respiration
Temperature dependence
Q10
Modeling
Copula
Temperature
Sensitivity

ABSTRACT

Soil CO₂ efflux represents a complex interplay of biological and physical processes that result in the production and transfer of CO2 from soils to the atmosphere. Temperature has been widely recognized as a critical factor regulating soil CO2 efflux and is commonly utilized in deterministic empirical models to predict this important flux for the carbon cycle. This study introduces the Bernstein copula-based cosimulation (BCC) as a data-driven probabilistic approach to model the temperature-soil CO2 efflux relationship. The BCC accounts for the joint probability distribution and temporal dependence of soil CO2 efflux, which are often overlooked in deterministic models. The BCC was implemented as a proof of concept using two years of data on soil CO2 efflux conditioned by soil temperature in a temperate forest. The BBC accurately reproduced the original probability distribution, temporal dependency, and temperature-soil CO2 efflux relationship. Our findings show that a deterministic method, such as the commonly employed exponential relationship between soil CO2 efflux and temperature, is limited for comprehensively capturing the intricate nature of the temperature-soil CO₂ efflux relationship. This is due to the confounding and interacting effects of environmental drivers beyond temperature, which are not fully accounted for in such a deterministic approach. Furthermore, the BCC revealed that the probability density between the joint cumulative probability of temperature and soil CO2 efflux is not constant, which raises the concern that deterministic approaches introduce incorrect assumptions for estimating temperature-soil CO2 relationship. In conclusion, we propose that probabilistic approaches hold promise for effectively depicting

E-mail address: rvargas@udel.edu (R. Vargas).

^{*} Corresponding author.

 $^{^{1}}$ Authors contributed equally to this work.

1. Introduction

Soil CO₂ efflux is a complex process that involves biological (e.g., plant and microbial respiration) and physical processes (e.g., diffusion, mass transport) that ultimately result in the production and transport of CO₂ from soils to the atmosphere (Raich and Schlesinger, 1992; Ryan and Law, 2005). This flux at an estimated 87 \pm 9 PgC y^{-1} is poorly constrained in the global terrestrial carbon cycle (Jian et al., 2022) and represents substantial uncertainty in ecosystem-process models (Luo et al., 2015; Sulman et al., 2018; Wei et al., 2022).

Temperature has been recognized as a major control for soil CO₂ efflux and is widely used to model this flux (Bahn et al., 2010; Phillips et al., 2017; Tuomi et al., 2008). Therefore, rising temperatures are expected to stimulate soil CO2 efflux by accelerating autotrophic respiration rates and heterotrophic decomposition of the organic matter (Bond-Lamberty et al., 2018). This expectation is based on observations (Bradford et al., 2008) that corroborate the generalities of the temperature dependence of biochemical processes described mathematically in the late 19th century (Arrhenius, 1889; Van't Hoff, 1899). However, the assumption by Arrhenius and van't Hoff of constant temperature sensitivity at all temperatures is incorrect to model soil CO2 efflux in mesic (Curiel Yuste et al., 2004; Lloyd and Taylor, 1994) and especially in water-limited ecosystems (Almagro et al., 2009; Leon et al., 2014). This limitation has motivated the development of other deterministic empirical models by incorporating temperature-moisture relationships (Del Grosso et al., 2005; Orchard and Cook, 1983; Vargas et al., 2010), mechanistic approaches to integrate temperature-sensitive enzymatic processes with constraints of substrate supply (Davidson et al., 2012; Todd-Brown et al., 2012), or macromolecular rate theory (Alster et al., 2020). Over time, the complexity of deterministic models for soil carbon dynamics and soil CO2 efflux has increased to capture the many biogeochemical mechanisms involved (Abramoff et al., 2022; Sulman et al., 2018).

An alternative approach for predicting soil CO₂ efflux is stochastic or probabilistic modeling. These approaches aim to understand a system's behavior regarding its underlying probability distribution and make independent predictions based on this information. Probabilistic models are assumed to improve our understanding of the complex dynamics of soil CO2 efflux and its complex response to multiple environmental factors (e.g., warming, precipitation events, and compounding effects). Several studies have incorporated these principles to model soil CO₂ efflux in Mediterranean (Anjileli et al., 2019) and temperate (Daly et al., 2008) ecosystems and at the global scale (Huang et al., 2021). Other studies have used these principles to predict and partition soil CO2 efflux (Zhou et al., 2010), and study the response to elevated CO₂ (Gao et al., 2020) or extreme heat events (Anjileli et al., 2021). Although there are fewer examples of probabilistic models when compared to deterministic models, there is an increasing interest in representing stochasticity for modeling soil organic matter dynamics (Waring et al., 2020).

We propose that the temporal variability of soil CO_2 efflux can be modeled based on its dependency on temperature using a probabilistic cosimulation approach. This study uses a copula approach based on probability theory (Sklar, 1959) to reproduce complex relationships between variables (Nelsen, 2006). Copulas are statistical tools that describe the dependence structure between random variables regardless of their marginal distributions. In a copula-based cosimulation, the marginal distribution function of each variable is modeled separately. Then, the copula linked to those individual distribution functions is modeled to capture their joint dependence.

For this study, we apply the Bernstein copula-based cosimulation (BCC) method as it does not assume linearity or Gaussian distribution,

which are characteristics of the relationship between temperature and soil CO_2 efflux (Vázquez-Ramírez et al., 2023). Briefly, the BCC is based on stochastic realizations obtained by the global optimization method (i. e., differential evolution) where the objective functions are univariate probability distribution, dependency coefficients, and a semivariogram model (Le et al., 2020; Vázquez-Ramírez et al., 2023). This approach can simulate the behavior of systems characterized by multiple stochastic interdependent processes while reproducing the statistical properties and temporal dependency. We postulate that the BCC is a superior approach to model soil CO_2 efflux because it incorporates information on the joint probability distribution and temporal dependence of this phenomenon that traditional deterministic methods ignore. To test this, the BCC was implemented as a proof of concept using information on soil CO_2 efflux and soil temperature from a temperate forest.

2. Methods

Two approaches for modeling soil CO_2 efflux based on soil temperature information are presented. The two approaches were implemented using dependency functions based on deterministic or probabilistic methods. The first approach is a deterministic method represented by the widely used exponential relationship between soil CO_2 efflux and soil temperature, represented by an exponential function-based estimation (EFE; Lloyd and Taylor, 1994). The second approach is a probabilistic method represented by the BCC conditioned by temperature (Le et al., 2020). These dependency functions based on soil temperature were evaluated using publicly available information on soil CO_2 efflux and temperature from a temperate forest (Vargas and Allen, 2008a, 2008b; Vargas et al., 2010).

2.1. Study site

The study site is located at the University of California James San Jacinto Mountain Reserve, a UC Natural Reserve System field station. The Reserve is a mixed conifer-oak forest at 1640 m.a.s.l. in the San Jacinto Mountains, CA, USA (33° 48′ 30″ N, 116° 46′ 40″ W). Most precipitation occurs between November and April, with a mean annual rainfall of 640 mm and a mean air temperature of 10.3 °C. Measurements were performed at an area with mature woody vegetation with prominent individuals (DBH > 30 cm) of *Quercus chrysolepis* Leibem. (Canyon live oak), and *Pinus ponderosa* C. Lawson (Ponderosa Pine). Soil and fine root characteristics have been described in previous studies (Vargas and Allen, 2008a, 2008b). Soil bulk density was 0.9 g/cm³, and soil texture was 83 % sand, 10 % silt, and 7 % clay. Total soil carbon and nitrogen (0–16 cm depth) were 3.1 % (±0.5 s.d.) and 0.08 % (±0.02 s.d.), respectively. Fine root biomass (0–16 cm) was 18 g m $^{-2}$, and fine root nitrogen was 0.58 % (±0.23 s.d.).

Soil CO_2 efflux measurements were performed using the flux gradient method described for the study site in multiple studies (Vargas et al., 2010; Vargas and Allen, 2008a, 2008b). Briefly, soil CO_2 concentrations (CARBOCAP, GMM 222, Vaisala, Helsinki, Finland), temperature, and moisture (Decagon, ECHO) were measured at three locations and three soil depths (2, 8, 16 cm). All measurements were recorded at 5-min intervals. Soil CO_2 efflux was calculated using Fick's first law of diffusion, and the diffusivity of soil CO_2 in the soil profile was estimated using the Moldrup model (Moldrup et al., 1999). Vargas and Allen (2008b) reported an r^2 of 0.94 between soil CO_2 efflux measured using a soil chamber (Li-8100-102) connected to a soil respiration system (LI-8100, LI-COR Lincoln, NE, USA) and the values calculated from the gradient method. The soil CO_2 efflux and temperature measurements used in this study are publicly available and have been incorporated into

the COSORE dataset (Bond-Lamberty et al., 2020). For the analyses presented in this study, data were aggregated into daily means of soil CO₂ efflux from January 1st 2006 to December 31st 2007.

2.2. Modeling soil CO2 efflux

The workflow for modeling soil CO_2 efflux included three steps: (a) Input data, (b) Dependency function modeling, and (c) Prediction and Uncertainty (Supplementary Fig. S1).

2.2.1. Input data

The input data consisted of two variables $X = \{x_1, x_2,, x_n\}$ and $Y = \{y_1, y_2,, y_n\}$ representing time series with the same length of data n; X is the time series of the independent variable (i.e., temperature) and Y is the time series of the dependent variable (i.e., soil CO_2 efflux). Then, 50 % of the data were used to train the models (i.e., training data) and the remaining data (i.e., test data) to validate the output of the models. Training data was selected from odd days (Fig. S2) and test data from even days (Fig. S3) to preserve the original time series' statistical properties and temporal dependence. Therefore, the performance of model outputs was tested on how well they represent the statistical properties and temporal dependence exhibited in the test dataset (Vargas and Le, 2023).

2.2.2. Dependency function modeling

The deterministic approach (EFE) was represented by a widely used exponential function $Y=a+e^{bX}$. This deterministic function was parameterized using the training data of temperature (X) and soil CO₂ efflux (Y), where a and b are the parameters to obtain (Lloyd and Taylor, 1994). This equation is rooted in the first principles of thermodynamics and is widely used for modeling soil CO₂ efflux because of its simplicity and broad applicability (Tuomi et al., 2008). We applied the lm and predict functions of the stats package in R (R Code Team, 2022), and the parameter interval = "predict" in the function predict to determine a 95 % confidence interval range to represent the uncertainty of the model and the model output (i.e., prediction).

The probabilistic approach (BCC) consisted of: a) modeling the joint probability distribution through univariate probability distributions and the copula distribution approximated by the Bernstein polynomial and the Bernstein copula; and b) modeling the temporal dependency function of the variable of interest Y using the semivariogram function (Le et al., 2020). According to Sklar's theorem (Sklar, 1959), the joint probability distribution can be decomposed into the univariate probability functions F(x), G(y), and the copula function C: H(x,y) =C(F(x), G(y)). This means that rather than modeling the joint probability function directly from a scatterplot representing the data of X and Y, its decomposition is modeled. To do this, the cumulative univariate distributions $F_n(x)$ and $G_n(y)$ and the empirical copula (i.e., Pseudoobservation) $C_n(F_n(x), G_n(y))$ were calculated. The distributions $F_n(x)$ and $G_n(y)$ were modeled by the Bernstein polynomial, and the distribution $C_n(F_n(x), G_n(y))$ was approximated by the Bernstein copula (Le et al., 2020). Finally, the empirical variogram was calculated from the soil CO2 efflux values using the training data, and a semivariogram model (e.g., spherical, exponential, or Gaussian semivariogram) was fitted to preserve the information on the temporal dependency.

2.2.3. Prediction & uncertainty

Soil CO_2 efflux conditioned by the temperature was predicted based on the two dependency functions (i.e., deterministic and probabilistic) described in the previous section. Both approaches were parameterized with the training data, and test data was used for prediction purposes. We calculated uncertainty for EFE by fitting the temperature of the test data to predict soil CO_2 efflux and its corresponding 95 % confidence interval range. For the BCC method, one hundred simulations of soil CO_2 efflux using the temperature of the test data were predicted to calculate

the probability and uncertainty (Le et al., 2020). Then, a global optimization method (i.e., differential evolution) was applied to obtain the soil CO_2 efflux prediction with similar univariate behavior and temporal and joint dependence as the training data. The 95 % quartile range was predicted for the one hundred simulations to determine BCC's uncertainty range. Finally, the predictions and uncertainties for EFE and BCC were compared against the soil CO_2 efflux values from the test data. The algorithm is implemented using the RGEOSTAD tools (Díaz-Viera et al., 2021) and is coded using the R software (R Code Team, 2022).

3. Results

3.1. Input data

We analyzed two years of information on daily temperature [°C] and soil CO_2 efflux [µmol m $^{-2}$ s $^{-1}$] (Fig. 1a and b). The statistical properties of the complete dataset of temperature were: mean 10.9 °C, median 9.5 °C, standard deviation 8.4 °C, minimum -1.8 °C, and maximum 28.7 °C (Fig. 1a). The statistical properties of the temperature training data were: mean 10.9 °C, median 9.5 °C, standard deviation 8.5 °C, minimum -1.4 °C, and maximum 28.7 °C (Fig. S2a). The statistical properties of the temperature test data were: mean 10.9 °C, median 9.6 °C, standard deviation 8.4 °C, minimum -1.8 °C, and maximum 28.4 °C (Fig. S3a). No statistically significant differences existed between the complete dataset, training data, or test data.

The statistical properties of the complete dataset of soil CO $_2$ efflux were: mean 1.1 $\mu mol~m^{-2}~s^{-1}$, median 0.9 $\mu mol~m^{-2}~s^{-1}$, standard deviation 0.6 $\mu mol~m^{-2}~s^{-1}$, minimum 0.2 $\mu mol~m^{-2}~s^{-1}$, and maximum 3.9 $\mu mol~m^{-2}~s^{-1}$ (Fig. 1b). The statistical properties of the soil CO $_2$ efflux training data were: mean 1.1 $\mu mol~m^{-2}~s^{-1}$, median 0.9 $\mu mol~m^{-2}~s^{-1}$, standard deviation 0.6 $\mu mol~m^{-2}~s^{-1}$, minimum 0.2 $\mu mol~m^{-2}~s^{-1}$, and maximum 3.2 $\mu mol~m^{-2}~s^{-1}$ (Fig. S2b). The statistical properties of the soil CO $_2$ efflux test data were: mean 1.1 $\mu mol~m^{-2}~s^{-1}$, median 0.9 $\mu mol~m^{-2}~s^{-1}$, standard deviation 0.6 $\mu mol~m^{-2}~s^{-1}$, minimum 0.2 $\mu mol~m^{-2}~s^{-1}$, and maximum 3.9 $\mu mol~m^{-2}~s^{-1}$ (Fig. S3b). Similarly, there were no statistically significant differences between the complete dataset, training data, or test data.

The following statistics represent the relationship between soil CO_2 efflux and temperature: Pearson = 0.69, Spearman = 0.71, Kendall = 0.52 (Fig. 1c). These results are identical for the complete dataset, training, and test data. An experimental semivariogram was used to explore and represented the temporal dependency of soil CO_2 efflux. In all cases (i.e., complete dataset, training, and test data), the sum of the points in the experimental semivariogram was equal to 0.5, demonstrating that the temporal dependencies of the time series were similar (Fig. S4a, b, and c).

3.2. Dependency function modeling

3.2.1. Using a deterministic approach

An exponential function (*EFE method*) was fitted using the training data obtaining the following equation: CO_2 *efflux* = 0.598 + $e^{0.044*Temperature}$ with an adjusted R^2 = 0.5 and a root mean square error (RMSE) of 0.36 (Fig. 2). In addition, a 95 % confidence interval was estimated for this model, which was then used to quantify the prediction uncertainty (Fig. 2).

3.2.2. Using a probabilistic approach

The first step to model soil CO_2 efflux with the Bernstein copula (BCC method) was to decompose the empirical bivariate joint probability distribution (Fig. 3a) into the empirical univariate probability distributions (black dots in Fig. 3b and c) and the empirical copula function (black dots in Fig. 3d). Then, the empirical univariate distributions were approximated by the Bernstein polynomial (blue line in Fig. 3b and c), and the empirical copula function was approximated using the Bernstein copula (blue lines in Fig. 3d). The cumulative probability distribution of

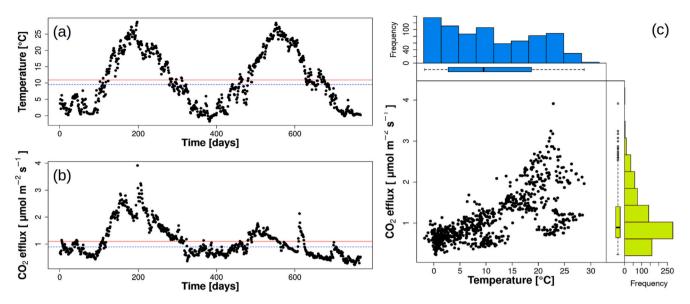


Fig. 1. Daily time series of (a) temperature, (b) soil CO₂ efflux, and the (c) bivariate relationship between them. The mean is represented in red, and the median in blue in panels (a) and (b). Panel (c) includes the univariate marginal distribution of each variable.

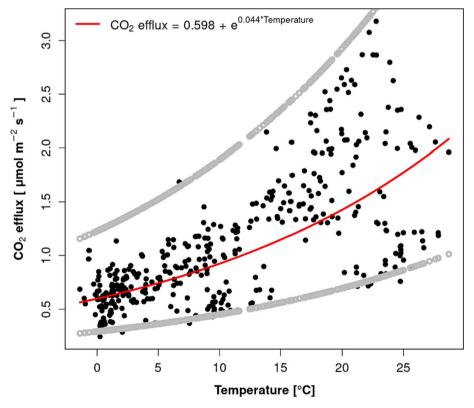


Fig. 2. Deterministic dependency model based on the exponential relationship between temperature and soil CO₂ efflux. The red line was generated by fitting an exponential model using the training data, and the gray dots represent the 95 % confidence intervals.

soil CO₂ efflux conditioned by temperature was not constant (Fig. 3d); where at a given soil temperature, a range of different soil CO₂ efflux values was possible, and these ranges varied across temperatures. To illustrate this, we show the cumulative probability distribution of soil CO₂ efflux for categories of soil temperature from 0 to 25 °C (Fig. 4). These results emphasize that the magnitude and distribution of soil CO₂ efflux were variable across different soil temperature ranges. In contrast, the EFE method predicted a deterministic value of soil CO₂ efflux for a given range of temperatures (Fig. 2).

Once the empirical univariate probability distribution and the empirical copula functions were modeled, the joint probability distribution of soil temperature and soil CO_2 efflux was fully constructed. The Bernstein polynomials and the Bernstein copula approximate empirical marginal functions and empirical copulas very well (Fig. 3b, c, and d). Therefore, the joint distribution function between temperature and soil CO_2 efflux was fully constructed. According to Sklar's theorem (Sklar, 1959), the dependency relationship between the variables X and Y is entirely contained by the copula function $C_n(F_n(x), G_n(y))$ (Fig. 3d) and

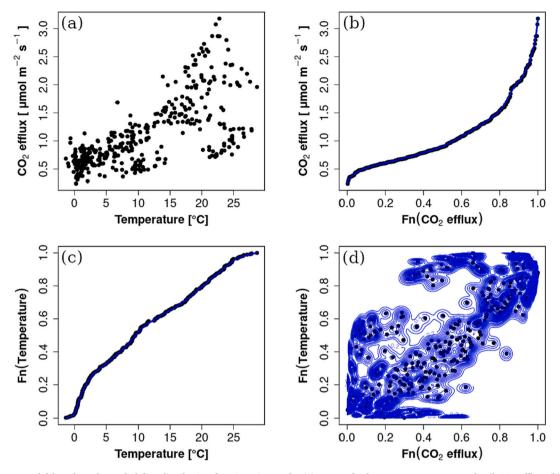


Fig. 3. Dependency model based on the probability distribution function via copula: (a) scatterplot between temperature and soil CO_2 efflux; (b) univariate distribution function of soil CO_2 efflux approximated by the Bernstein polynomial (blue line); (c) univariate distribution function of temperature approximated by the Bernstein polynomial (blue line); and (d) copula function approximated by Bernstein copula (blue contour) in copula space (pseudo-observation).

this dependency does not depend on the marginal functions $F_n(x)$, $G_n(y)$ (Fig. 3b and c).

Finally, a semivariogram model was fitted to the experimental semivariogram for the training data of soil CO_2 efflux where the sum of the points was equal to 0.5 (Fig. S4b). A spherical variogram model was obtained with a nugget of 0.003, a sill of 0.075, and a temporal correlation range of 24 days (Fig. S5). This fitted variogram model had a sum of squared error (SSE) of 0.002 compared to the experimental semi-variogram from the training dataset.

3.3. Prediction & uncertainty

3.3.1. Comparison of univariate statistical properties

The statistical properties of the predicted soil CO $_2$ efflux using the deterministic approach (i.e., *EFE*) were: mean 1.0 µmol m $^{-2}$ s $^{-1}$, median 0.9 µmol m $^{-2}$ s $^{-1}$, standard deviation 0.4 µmol m $^{-2}$ s $^{-1}$, minimum 0.6 µmol m $^{-2}$ s $^{-1}$, and maximum 2.1 µmol m $^{-2}$ s $^{-1}$ (red dots in Fig. 5a). The residuals show that the deterministic model underestimated soil CO $_2$ efflux during the growing season of the first year and overestimated soil CO $_2$ efflux for most of the mid-to-late part of the second year (Fig. S6a). The statistical properties of the predicted soil CO $_2$ efflux using the probabilistic approach (i.e., *BCC*) were: mean 1.1 µmol m $^{-2}$ s $^{-1}$, median 0.9 µmol m $^{-2}$ s $^{-1}$, standard deviation 0.6 µmol m $^{-2}$ s $^{-1}$, minimum 0.3 µmol m $^{-2}$ s $^{-1}$, and maximum 2.9 µmol m $^{-2}$ s $^{-1}$ (blue dots in Fig. 5b). The residuals show that the probabilistic approach properly represents the magnitudes and patterns of soil CO $_2$ efflux (Fig. S6b). Furthermore, the *p*-values of the Kolmogorov-Smirnov test for the comparison of the predictions with the test data (black dots in Fig. S7 and histogram-

boxplot in Fig. S8a) were 0 for the EFE (red triangle in Fig. S7 and histogram-boxplot in Fig. S8b) and 1 for the BCC (blue squares in Fig. S7 and histogram-boxplot in Fig. S8c). This shows that the probability distribution of the BCC prediction matches that of the test data, while the probability distribution of the EFE prediction differs significantly from the test data.

3.3.2. Comparison of dependency relationship

The dependency relationship between predicted soil CO_2 efflux and temperature (using the EFE) was represented by Pearson = 0.99, Spearman = 1, and Kendall = 1; suggesting a spurious relationship as a result of the deterministic nature of the approach (Fig. 5c). The dependency relationship between predicted soil CO_2 efflux and temperature (using the BCC) was represented by Pearson = 0.7, Spearman = 0.71, and Kendall = 0.51; preserving the dependency relationship and scatter plot pattern of the complete dataset, training, or test data (Fig. 5d).

3.3.3. Comparison of temporal dependence or variability

The sum of the points of the experimental semivariogram for the predicted soil CO_2 efflux using the EFE was 0.1, demonstrating a different temporal dependency than the complete dataset or the test data (Fig. S9). The error or difference between the empirical variogram of the prediction of the EFE with the test data was 0.4 (Fig. S9). The sum of the points of the experimental semivariogram for the predicted soil CO_2 efflux using the BCC was 0.5. The error or difference between the empirical variogram of the prediction of the BCC with the test data was 0.05 and 8 times smaller than the error from the EFE (Fig. S9).

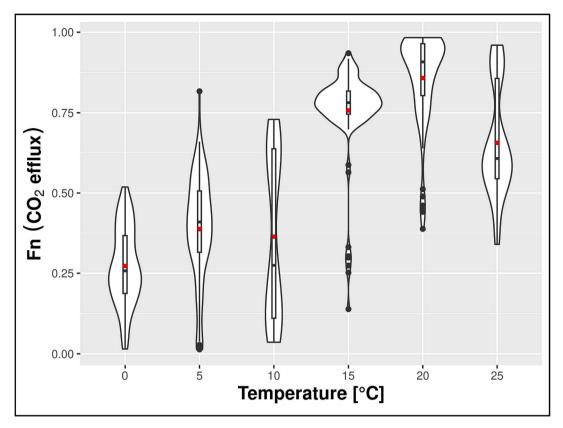


Fig. 4. Violin plots between the conditioned temperature and the cumulative probability distribution of soil CO₂ efflux predicted using the BCC. This figure demonstrates that the probability of the soil CO₂ efflux value is not constant for a given temperature and complements the information presented in Fig. 3b, c, and d.

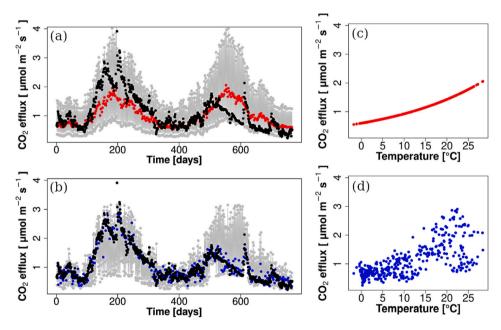


Fig. 5. Temporal distribution of CO_2 efflux prediction conditioned by temperature using: (a) an EFE as a deterministic approach (red dots), and (b) BCC as a stochastic approach (blue dots). The black dots represent the independent test data and the gray dots in both the associated uncertainty for each one of the approaches. Scatterplots between temperature (from test data) and the soil CO_2 efflux predictions from EFE (c) and BCC (d).

3.3.4. Comparison of uncertainties and errors

From the models built using the training data, the predictions of the soil $\rm CO_2$ efflux conditioned by the temperature of the test data (red dots in Fig. 5a for the EFE and blue dots in Fig. 5b for the BCC) and its associated uncertainty range were made (the gray area in Fig. 5a for the

EFE and the gray area in Fig. 5b for the BCC). The temporal distribution pattern of the EFE prediction points is different from soil CO_2 efflux values from the test data (Fig. 5a). In contrast, the pattern of BCC prediction points is closer to soil CO_2 efflux values from the test data (Fig. 5b). The prediction of soil CO_2 efflux using the BCC method

presents a total error of 7.61 % when compared to the test data (i.e., the absolute sum of the blue dots in Fig. S6 divided by the total sum of soil $\rm CO_2$ efflux from the test data), while the prediction using the EFE method has a 15.22 % (i.e., the absolute sum of the red dots in Fig. S6 divided by the total sum of soil $\rm CO_2$ efflux from the test data), or about a 50 % larger error than the BCC method. The total uncertainty of the EFE method is about 1.4 times more than the BCC method (i.e., the sum of the uncertainty range in Figs. S6a and S6b). The results show that the errors and uncertainty ranges of the BCC method are substantially lower than those of the EFE method (Fig. S6).

4. Discussion

Our results support the expectation that BCC is a superior approach to model soil CO2 efflux only using temperature information than a simple deterministic approach (i.e., EFE method). This commonly used deterministic approach provides an apparent temperature-soil CO2 efflux response that does not accurately represent the actual dependency relationship between these variables. The dependency relationship between soil CO2 efflux and temperature has commonly been used to derive empirical response functions, estimate the temperature dependency of soil CO2 efflux (i.e., Q10), and derive annual estimates of soil CO₂ efflux based on sporadic measurements in time (Bahn et al., 2010; Davidson et al., 2006). Although temperature is an essential environmental driver for many underlying biophysical processes that regulate soil CO2 efflux, its direct influence on this flux can only be approximated by empirical models. This is because there are multiple challenges to identifying the actual temperature-soil CO2 efflux dependency relationship that includes belowground carbon allocation in plants (Chen et al., 2011; Meeran et al., 2021), the influence of water on soil CO2 diffusivity and rewetting of soils (Kim et al., 2012), or thermal gradients in the soil (Graf et al., 2008; Tang et al., 2005; Vargas and Allen, 2008c). It has been proposed that because of these challenges, deterministic temperature response functions are limited in representing the temporal variability of soil CO₂ efflux and predicting the effects of climate change (Sierra, 2012; Subke and Bahn, 2010). Although we advocate for increased complexity in models to capture the many biogeochemical mechanisms involved (Abramoff et al., 2022; Sulman et al., 2018), we propose that a shift from a deterministic to a probabilistic approach should be explored for improving dependency relationships in simple empirical modeling approaches.

This study demonstrates the limitations of an empirical deterministic approach exemplified by a widely used exponential function (i.e., EFE) to represent the apparent temperature-soil CO₂ relationship and predict how soil CO2 fluxes change over time. This empirical approach attempts to describe the pairwise dependence of these variables but assumes that the same univariate marginal distribution characterizes their behavior. Our results show this is not the case (Fig. 1c); consequently, the EFE violates this assumption and produces a deterministic output with limitations to predict soil CO2 efflux (Fig. 5). The EFE model output generates a very strong spurious relationship between these variables that is not observed with the original data (Fig. 5c). A traditional explanation has been that a simple empirical model representing the apparent temperature-soil CO2 relationship may not be accurate because there are confounding factors such as water pulses and changes in substrate supply from plants to roots and mycorrhizae (Vargas et al., 2010; Vargas and Allen, 2008a). The scientific literature has many examples of why confounding factors (e.g., light, water, photosynthesis) challenge the deterministic modeling of soil CO2 fluxes only based on temperature (Davidson et al., 1998; Liu et al., 2006; Zhang et al., 2018), but this simple deterministic approach is constantly used and usually overrepresents the influence of temperature in predicting terrestrial respiration (Davidson et al., 2006; Murray-Tortarolo et al., 2022; Wei et al., 2022).

We propose that the BCC is a promising approach to modeling the actual temperature-soil ${\rm CO_2}$ relationship by combining information

from the joint cumulative probability distribution function using a copula with input from their temporal dependency. Our results demonstrate that the joint cumulative probability distributions were accurately approximated by the Bernstein polynomial and the Bernstein copula (Fig. 3). The Bernstein copula showed a positive relationship between the cumulative probability of soil $\rm CO_2$ efflux and the cumulative probability of soil temperature (Fig. 3d) and preserved the original characteristics of the relationship (Figs. 2 and 5d). We highlight that the BCC can represent non-linear and non-Gaussian relationships between variables, which are more representative of real-world data (e.g., the relationship between soil $\rm CO_2$ efflux and temperature).

We found that the probability density between the joint cumulative probability of temperature and soil CO2 efflux is not constant (Fig. 3d and Fig. 4). These results raise questions about deterministic approaches that predict specific soil CO₂ efflux rates given specific soil temperatures (Fig. 5c) and provide support for the concern of potential incorrect estimations of apparent temperature-soil CO2 relationships (Davidson et al., 2006; Subke and Bahn, 2010). The BCC accurately reproduced the temperature-soil CO2 relationship by preserving the variability in their joint cumulative probability distributions (Fig. 5d). We postulate that the information on the joint probability distribution provides insights about ranges of temperature and soil CO2 efflux of interest that can be explored with manipulative experiments to understand the biophysical processes that regulate this variability. For example, temperatures between 5 and 10° C showed the largest variability in the cumulative probability distribution of soil CO2 efflux, while there were more constrained values at temperatures between 15 and 20° C (Fig. 4). These results highlight the importance of understanding the joint probability distribution to describe dependence relationships and may prove relevant to quantify and improve theoretical approaches such as the thermal adaptation of microbial respiration to changes in temperature (Alster et al., 2020; Bradford et al., 2008).

The BCC was able to properly show how soil CO_2 efflux varies in time based only on temperature information. Our results show that the univariate cumulative probability distribution of predicted soil CO2 efflux using BCC was similar to the one of the test data (Fig. S7 and Fig. S8) and is also capable of representing the temporal dependence (Fig. S9). In contrast, the EFE could not describe the univariate cumulative probability distribution (Fig. S7 and Fig. S8) nor the temporal dependence of soil CO₂ efflux (Fig. S9). Reproducing the temporal dependence in time series is essential to preserve the sequential information present in the data. The BCC has an additional advantage as it captures patterns and dependencies in the time series and the dynamics and relationships between data points in time. Arguably, a model can reproduce the mean and the probability distribution of the data but not necessarily the temporal dependencies that provide interpretability of the model output. Previous studies have analyzed the importance of evaluating model outputs in the frequency domain to understand data-model disagreements and provide insights into biophysical processes or environmental events that may challenge model performance (Dietze et al., 2011; Stoy et al., 2013; Vargas et al., 2010, 2018). Replicating temporal dependence is a crucial aspect that is critical for constructing better models, obtaining interpretability, and extracting relevant features from time-varying data (e.g., seasonality, pulses). Overall, it enables us to capture the data's sequential pattern, and the BCC effectively uses it with potential applications to model data relationships across diverse scientific domains (Le et al., 2020; Vargas and Le, 2023).

The BCC approach has limitations and future potential. Its complexity and the need for robust datasets for model training can pose challenges, particularly regarding data availability and the computational resources required for global optimization methods. Moreover, while the BCC offers improved precision over deterministic models, its predictive power is still contingent on the quality and representativeness of the input data. The BCC approach has substantially higher computational costs than the EFE for model construction and prediction. Nevertheless, this computational cost can be overcome using

representative subsampling techniques like the temporal Latin Hypercube Sampling (Vargas and Le, 2023) combined with parallelization in multiple processors, depending on the length of the training data. Furthermore, the BBC approach should be tested with longer time series, and across various ecosystems and weather conditions. This will help evaluate its effectiveness across biomes and non-stationary conditions such as precipitation pulses, which have a tendency to generate hotmoments for greenhouse gas fluxes (Kim et al., 2012; Petrakis et al., 2017). We highlight that the BCC approach can be extended to multivariate cases with several environmental variables to predict a set of greenhouse gasses of interest, including CO₂, CH₄, and N₂O efflux. This extension can be achieved using the vine copula (Czado and Nagler, 2022) and can be implemented to represent the spatial variability of soil gas fluxes.

5. Conclusion

We conclude that the probabilistic method based on the Bernstein copula (i.e., BCC) produces better precision and accuracy than the deterministic method based on a commonly used exponential function (i.e., EFE). The relevance of the BCC approach is that the dependency relationship between soil CO2 efflux and temperature is transformed from the time domain to the probability domain, and this dependency relationship is then modeled using the copula function and separated from the univariate probability distributions. We highlight that soil CO2 efflux and temperature have different univariate marginal distributions and, therefore, question deterministic approaches that assume the same univariate distribution of two variables when analyzing pairwise dependencies. Furthermore, the BCC reproduces the temporal dependence function of soil CO₂ efflux (by including semivariogram information), a characteristic often neglected by most modeling approaches, assuming that the temporal variability is inherently described by analyzing pairwise dependencies. Ultimately, probabilistic methods have the potential to accurately delineate dependency connections for soil CO2 efflux modeling and in enhancing projections of the consequences of weather variability and climate change.

CRediT authorship contribution statement

Van Huong Le: Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. Rodrigo Vargas: Conceptualization, Data curation, Investigation, Methodology, Resources, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data is available online at Figshare: https://doi.org/10.6084/m9.figshare.11739459.v1

Acknowledgments

This study was supported by NASA (80NSSC21K171520) and the Department of Energy (DE-SC0023099) of the United States.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi. org/10.1016/j.scitotenv.2023.169391.

References

- Abramoff, R.Z., Guenet, B., Zhang, H., Georgiou, K., Xu, X., Viscarra Rossel, R.A., Yuan, W., Ciais, P., 2022. Improved global-scale predictions of soil carbon stocks with Millennial Version 2. Soil Biol. Biochem. 164, 108466.
- Almagro, M., López, J., Querejeta, J.I., Martínez-Mena, M., 2009. Temperature dependence of soil CO₂ efflux is strongly modulated by seasonal patterns of moisture availability in a Mediterranean ecosystem. Soil Biol. Biochem. 41 (3), 594–605.
- Alster, C.J., von Fischer, J.C., Allison, S.D., Treseder, K.K., 2020. Embracing a new paradigm for temperature sensitivity of soil microbes. Glob. Chang. Biol. 26 (6), 3221–3229.
- Anjileli, H., Moftakhari, H.R., Mazdiyasni, O., Norouzi, H., Ashraf, S., Farahmand, A., Bowler, P., Azarderakhsh, M., Huxman, T.E., AghaKouchak, A., 2019. Analyzing high-frequency soil respiration using a probabilistic model in a semiarid, Mediterranean climate. J. Geophys. Res. Biogeosci. 124 (3), 509–520.
- Anjileli, H., Huning, L.S., Moftakhari, H., Ashraf, S., Asanjan, A.A., Norouzi, H., AghaKouchak, A., 2021. Extreme heat events heighten soil respiration. Sci. Rep. 11 (1), 6632.
- Arrhenius, S., 1889. Über die Reaktionsgeschwindigkeit bei der Inversion von Rohrzucker durch Säuren. Z. Phys. Chem. 4U (1), 226–248.
- Bahn, M., Reichstein, M., Davidson, E.A., Gruenzweig, J., Jung, M., Carbone, M.S., Epron, D., Misson, L., Nouvellon, Y., Roupsard, O., Savage, K., Trumbore, S.E., Gimeno, C., Curiel Yuste, J., Tang, J., Vargas, R., Janssens, I.A., 2010. Soil respiration at mean annual temperature predicts annual total across vegetation types and biomes. Biogeosciences 7 (7), 2147–2157.
- Bond-Lamberty, B., Bailey, V.L., Chen, M., Gough, C.M., Vargas, R., 2018. Globally rising soil heterotrophic respiration over recent decades. Nature 560, 80–83.
- Bond-Lamberty, B., Christianson, D.S., Malhotra, A., Pennington, S.C., Sihi, D., AghaKouchak, A., Anjileli, H., Altaf Arain, M., Armesto, J.J., Ashraf, S., Ataka, M., Baldocchi, D., Andrew Black, T., Buchmann, N., Carbone, M.S., Chang, S., Crill, P., Curtis, P.S., Davidson, E.A., Zou, J., 2020. COSORE: A community database for continuous soil respiration and other soil-atmosphere greenhouse gas flux data. Glob. Chang. Biol. 249, 434.
- Bradford, M.A., Davies, C.A., Frey, S.D., Maddox, T.R., Melillo, J.M., Mohan, J.E., Reynolds, J.F., Treseder, K.K., Wallenstein, M.D., 2008. Thermal adaptation of soil microbial respiration to elevated temperature. Ecol. Lett. 11 (12), 1316–1327.
- Chen, G.-S., Yang, Y.-S., Guo, J.-F., Xie, J.-S., Yang, Z.-J., 2011. Relationships between carbon allocation and partitioning of soil respiration across world mature forests. Plant Ecol. 212 (2), 195–206.
- Curiel Yuste, J., Janssens, I.A., Carrara, A., Ceulemans, R., 2004. Annual Q(10) of soil respiration reflects plant phenological patterns as well as temperature sensitivity. Glob. Chang. Biol. 10 (2), 161–169.
- Czado, C., Nagler, T., 2022. Vine copula based modeling. Annu. Rev. Stat. Appl. 9 (1), 453–477.
- Daly, E., Oishi, A.C., Porporato, A., Katul, G.G., 2008. A stochastic model for daily subsurface CO₂ concentration and related soil respiration. Adv. Water Resour. 31, 987–994
- Davidson, E.A., Belk, E., Boone, R.D., 1998. Soil water content and temperature as independent or confounded factors controlling soil respiration in a temperate mixed hardwood forest. Glob. Chang. Biol. 4 (2), 217–227.
- Davidson, E.A., Janssens, I.A., Luo, Y.Q., 2006. On the variability of respiration in terrestrial ecosystems: moving beyond Q(10). Glob. Chang. Biol. 12 (2), 154–164.
- Davidson, E.A., Samanta, S., Caramori, S.S., Savage, K., 2012. The dual Arrhenius and Michaelis-Menten kinetics model for decomposition of soil organic matter at hourly to seasonal time scales. Glob. Chang. Biol. 18 (1), 371–384.
- Del Grosso, S.J., Parton, W.J., Mosier, A.R., Holland, E.A., Pendall, E., Schimel, D.S., Ojima, D.S., 2005. Modeling soil CO₂ emissions from ecosystems. Biogeochemistry 73 (1), 71–91.
- Díaz-Viera, M.A., Hernández-Maldonado, V., Méndez-Venegas, J., Mendoza-Torres, F., Le, V.H., Vázquez-Ramírez, D., 2021. RGEOSTAD: Un programa de código abierto para aplicaciones geoestadísticas basado en r-project.
- Dietze, M.C., Vargas, R., Richardson, A.D., 2011. Characterizing the performance of ecosystem models across time scales: A spectral analysis of the North American Carbon Program site-level synthesis. J. Geophys. Res. Biogeosci. 116 (G04029), 1–15
- Gao, Q., Wang, G., Xue, K., Yang, Y., Xie, J., Yu, H., Bai, S., Liu, F., He, Z., Ning, D., Hobbie, S.E., Reich, P.B., Zhou, J., 2020. Stimulation of soil respiration by elevated CO₂ is enhanced under nitrogen limitation in a decade-long grassland study. Proc. Natl. Acad. Sci. U. S. A. 117 (52), 33317–33324.
- Graf, A., Weihermüller, L., Huisman, J.A., Herbst, M., Bauer, J., Vereecken, H., 2008. Measurement depth effects on the apparent temperature sensitivity of soil respiration in field studies. Biogeosciences 5 (4), 1175–1188.
- Huang, H., Calabrese, S., Rodriguez-Iturbe, I., 2021. Variability of ecosystem carbon source from microbial respiration is controlled by rainfall dynamics. Proc. Natl. Acad. Sci. U. S. A. 118 (52) https://doi.org/10.1073/pnas.2115283118.
- Jian, J., Bailey, V., Dorheim, K., Konings, A.G., Hao, D., Shiklomanov, A.N., Snyder, A., Steele, M., Teramoto, M., Vargas, R., Bond-Lamberty, B., 2022. Historically inconsistent productivity and respiration fluxes in the global terrestrial carbon cycle. Nat. Commun. 13 (1), 1733.
- Kim, D.G., Vargas, R., Bond-Lamberty, B., Turetsky, M.R., 2012. Effects of soil rewetting and thawing on soil gas fluxes: a review of current literature and suggestions for future research. Biogeosciences 9 (7), 2459–2483.
- Le, V.H., Díaz-Viera, M.A., Vázquez-Ramírez, D., del Valle-García, R., Erdely, A., Grana, D., 2020. Bernstein copula-based spatial cosimulation for petrophysical property prediction conditioned to elastic attributes. J. Pet. Sci. Eng. 193 (107382), 107382.

- Leon, E., Vargas, R., Bullock, S., Lopez, E., Panosso, A.R., La Scala, N., Jr., 2014. Hot spots, hot moments, and spatio-temporal controls on soil CO₂ efflux in a waterlimited ecosystem. Soil Biol. Biochem. 77, 12–21.
- Liu, Q., Edwards, N.T., Post, W.M., Gu, L., Ledford, J., Lenhart, S., 2006. Temperature-independent diel variation in soil respiration observed from a temperate deciduous forest. Glob. Chang. Biol. 12 (11), 2136–2145.
- Lloyd, J., Taylor, J.A., 1994. On the temperature dependence of soil respiration. Funct. Ecol. 8 (3), 315–323.
- Luo, Y., Keenan, T.F., Smith, M., 2015. Predictability of the terrestrial carbon cycle. Glob. Chang. Biol. 21 (5), 1737–1751.
- Meeran, K., Ingrisch, J., Reinthaler, D., Canarini, A., Müller, L., Pötsch, E.M., Richter, A., Wanek, W., Bahn, M., 2021. Warming and elevated CO₂ intensify drought and recovery responses of grassland carbon allocation to soil respiration. Glob. Chang. Biol. 27 (14), 3230–3243.
- Moldrup, P., Olesen, T., Yamaguchi, T., Schjonning, P., Rolston, D.E., 1999. Modeling diffusion and reaction in soils. IX. The bucking-ham-Burdine-Campbell equation for gas diffusivity in undisturbed soil. Soil Sci. 164 (8), 542–551.
- Murray-Tortarolo, G., Poulter, B., Vargas, R., Hayes, D., Michalak, A.M., Williams, C., Windham-Myers, L., Wang, J.A., Wickland, K.P., Butman, D., Tian, H., Sitch, S., Friedlingstein, P., O'Sullivan, M., Briggs, P., Arora, V., Lombardozzi, D., Jain, A.K., Yuan, W., Kurz, W., 2022. A process-model perspective on recent changes in the carbon cycle of North America. J. Geophys. Res. Biogeosci. 127 (9), e2022JG006904.
- Nelsen, R.B., 2006. An Introduction to Copulas. Springer, New York.
- Orchard, V.A., Cook, F.J., 1983. Relationship between soil respiration and soil moisture. Soil Biol. Biochem. 15 (4), 447–453.
- Petrakis, S., Seyfferth, A., Kan, J., Inamdar, S., Vargas, R., 2017. Influence of experimental extreme water pulses on greenhouse gas emissions from soils. Biogeochemistry 133, 147–164.
- Phillips, C.L., Bond-Lamberty, B., Desai, A.R., Lavoie, M., Risk, D., Tang, J.W., Todd-Brown, K., Vargas, R., 2017. The value of soil respiration measurements for interpreting and modeling terrestrial carbon cycling. Plant Soil 413 (1–2), 1–25.
- R Code Team, 2022. R: A Language and Environment for Statistical Computing. R
 Foundation for Statistical Computing, Vienna, Austria.
- Raich, J.W., Schlesinger, W.H., 1992. The global carbon dioxide flux in soil respiration and its relationship to vegetation and climate. Tellus B: Chem. Phys. Meteorol. 44 (2), 81–99.
- Ryan, M.G., Law, B.E., 2005. Interpreting, measuring, and modeling soil respiration. Biogeochemistry 73 (1), 3–27.
- Sierra, C.A., 2012. Temperature sensitivity of organic matter decomposition in the Arrhenius equation; some theoretical considerations. Biogeochemistry 108 (1), 1–15.
- Sklar, M., 1959. Fonctions de répartition à n dimensions et leurs marges. Annales de l'ISUP. 8. 229–231. https://hal.science/hal-04094463/document.
- Stoy, P.C., Dietze, M.C., Richardson, A.D., Vargas, R., Barr, A.G., Anderson, R.S., Arain, M.A., Baker, I.T., Black, T.A., Chen, J.M., Cook, R.B., Gough, C.M., Grant, R. F., Hollinger, D.Y., Izaurralde, R.C., Kucharik, C.J., Lafleur, P., Law, B.E., Liu, S., Weng, E., 2013. Evaluating the agreement between measurements and models of net ecosystem exchange at different times and timescales using wavelet coherence: an example using data from the North American Carbon Program Site-Level Interim Synthesis. Biogeosciences 10 (11), 6893–6909.
- Subke, J.-A., Bahn, M., 2010. On the 'temperature sensitivity' of soil respiration: can we use the immeasurable to predict the unknown? Soil Biol. Biochem. 42 (9), 1653-1656.

- Sulman, B.N., Moore, J.A.M., Abramoff, R., Averill, C., Kivlin, S., Georgiou, K., Sridhar, B., Hartman, M.D., Wang, G., Wieder, W.R., Bradford, M.A., Luo, Y., Mayes, M.A., Morrison, E., Riley, W.J., Salazar, A., Schimel, J.P., Tang, J., Classen, A. T., 2018. Multiple models and experiments underscore large uncertainty in soil carbon dynamics. Biogeochemistry 141 (2), 109–123.
- Tang, J., Baldocchi, D.D., Xu, L., 2005. Tree photosynthesis modulates soil respiration on a diurnal time scale. Glob. Chang. Biol. 11, 1298–1304.
- Todd-Brown, K.E.O., Hopkins, F.M., Kivlin, S.N., Talbot, J.M., Allison, S.D., 2012.
 A framework for representing microbial decomposition in coupled climate models.
 Biogeochemistry 109 (1), 19–33.
- Tuomi, M., Vanhala, P., Karhu, K., Fritze, H., Liski, J., 2008. Heterotrophic soil respiration—comparison of different models describing its temperature dependence. Ecol. Model. 211 (1), 182–190.
- Van't Hoff, J.H., 1899. Lectures on Theoretical and Physical Chemistry. Part 1.. Chemical Dynamics. Edward Arnold, London.
- Vargas, R., Allen, M.F., 2008a. Environmental controls and the influence of vegetation type, fine roots and rhizomorphs on diel and seasonal variation in soil respiration. New Phytol. 179 (2), 460–471.
- Vargas, R., Allen, M.F., 2008b. Dynamics of fine root, fungal rhizomorphs, and soil respiration in a mixed temperate forest: integrating sensors and observations. Vadose Zone J. 7 (3), 1055–1064.
- Vargas, R., Allen, M.F., 2008c. Diel patterns of soil respiration in a tropical forest after hurricane Wilma. J. Geophys. Res. 113 (G3), G03021.
- Vargas, R., Le, V.H., 2023. The paradox of assessing greenhouse gases from soils for nature-based solutions. Biogeosciences 20 (1), 15–26.
- Vargas, R., Detto, M., Baldocchi, D.D., Allen, M.F., 2010. Multiscale analysis of temporal variability of soil CO₂ production as influenced by weather and vegetation. Glob. Chang. Biol. 16 (5), 1589–1605.
- Vargas, R., Sánchez-Cañete, P.E., Serrano-Ortiz, P., Curiel Yuste, J., Domingo, F., López-Ballesteros, A., Oyonarte, C., 2018. Hot-moments of soil CO₂ efflux in a water-limited grassland. Soil Syst. 2 (3), 47.
- Vázquez-Ramírez, D., Van Huong, L., Díaz-Viera, M.A., del Valle-García, R., Erdely, A., 2023. Joint stochastic simulation of petrophysical properties with elastic attributes based on parametric copula models. Geofis. Int. 62 (2), 487–506.
- Waring, B.G., Sulman, B.N., Reed, S., Smith, A.P., Averill, C., Creamer, C.A., Cusack, D.F., Hall, S.J., Jastrow, J.D., Jilling, A., Kemner, K.M., Kleber, M., Liu, X.-J.A., Pett-Ridge, J., Schulz, M., 2020. From pools to flow: the PROMISE framework for new insights on soil carbon cycling in a changing world. Glob. Chang. Biol. 26 (12), 6631–6643.
- Wei, N., Xia, J., Zhou, J., Jiang, L., Cui, E., Ping, J., Luo, Y., 2022. Evolution of uncertainty in terrestrial carbon storage in earth system models from CMIP5 to CMIP6. J. Clim. 35 (17), 5483–5499.
- Zhang, Q., Phillips, R.P., Manzoni, S., Scott, R.L., Oishi, A.C., Finzi, A., Daly, E., Vargas, R., Novick, K.A., 2018. Changes in photosynthesis and soil moisture drive the seasonal soil respiration-temperature hysteresis relationship. Agric. For. Meteorol. 259, 184–195.
- Zhou, X.H., Luo, Y.Q., Gao, C., Verburg, P.S.J., Arnone, J.A., Darrouzet-Nardi, A., Schimel, D.S., 2010. Concurrent and lagged impacts of an anomalously warm year on autotrophic and heterotrophic components of soil respiration: a deconvolution analysis. New Phytol. 187 (1), 184–198.