Copula-based dependency model for CO_2 efflux prediction and its uncertainty quantification

Van Huong Le[1], Rodrigo Vargas[1]

[1]Department of Plant and Soil Sciences, University of Delaware, Newark, DE 19716, USA

email: vanle@udel.edu, rvargas@udel.edu

Abstract

A new data-driven method for the prediction of temporally distributed CO_2 efflux conditioned by environmental variables is presented. The method, namely Bernstein copula-based stochastic cosimulation [Le *et al.*, 2020], is based on Bernstein copula for the estimation of the joint probability distribution function and simulated annealing for the temporal distribution simulation. The proposed method can model complex non-linear relationships between variables in a fully nonparametric approach. The main advantage is that it does not require linear dependence between variables nor any distribution constraint.

This method is validated using a time series of CO_2 efflux conditioned to temperature in a temperate forest in Delaware. The results show that this method could reproduce the statistical properties and temporal relationships of the phenomena studied.

Introduction

Modeling the temporal distribution of CO_2 efflux is challenging due to the limited amount of data available and the uncertainty in measurements. For this reason, stochastic simulation approaches have been adopted to model the temporal distribution of the CO_2 efflux using environmental variables as secondary variables.

Goals

Propose and implement a copula-based dependency model that can correctly model the dependency relationships between variables that represent natural phenomena, and then make use of this model to predict variables of interest.

Brief introduction to copula

Let F_{XY} be a bivariate joint probability distribution function and F_X and F_Y be univariate (marginal) probability distribution functions. A copula is a function C: $[0,1]^2 \mapsto [0,1]$ such that for all x, y in $\mathbb{R}$,

$$F_{XY}(x,y) = C_{XY}\{F_X(x), F_Y(y)\} \tag{1}$$

If F_X and F_Y are continuous, then C_{XY} is unique; otherwise, C_{XY} is uniquely determined on $RanF_X \times RanF_Y$. Copula associated to a bivariate random vector (X,Y) describes the relationship between X and Y, and independently from their univariate probabilistic behavior [Sklar, 1959].

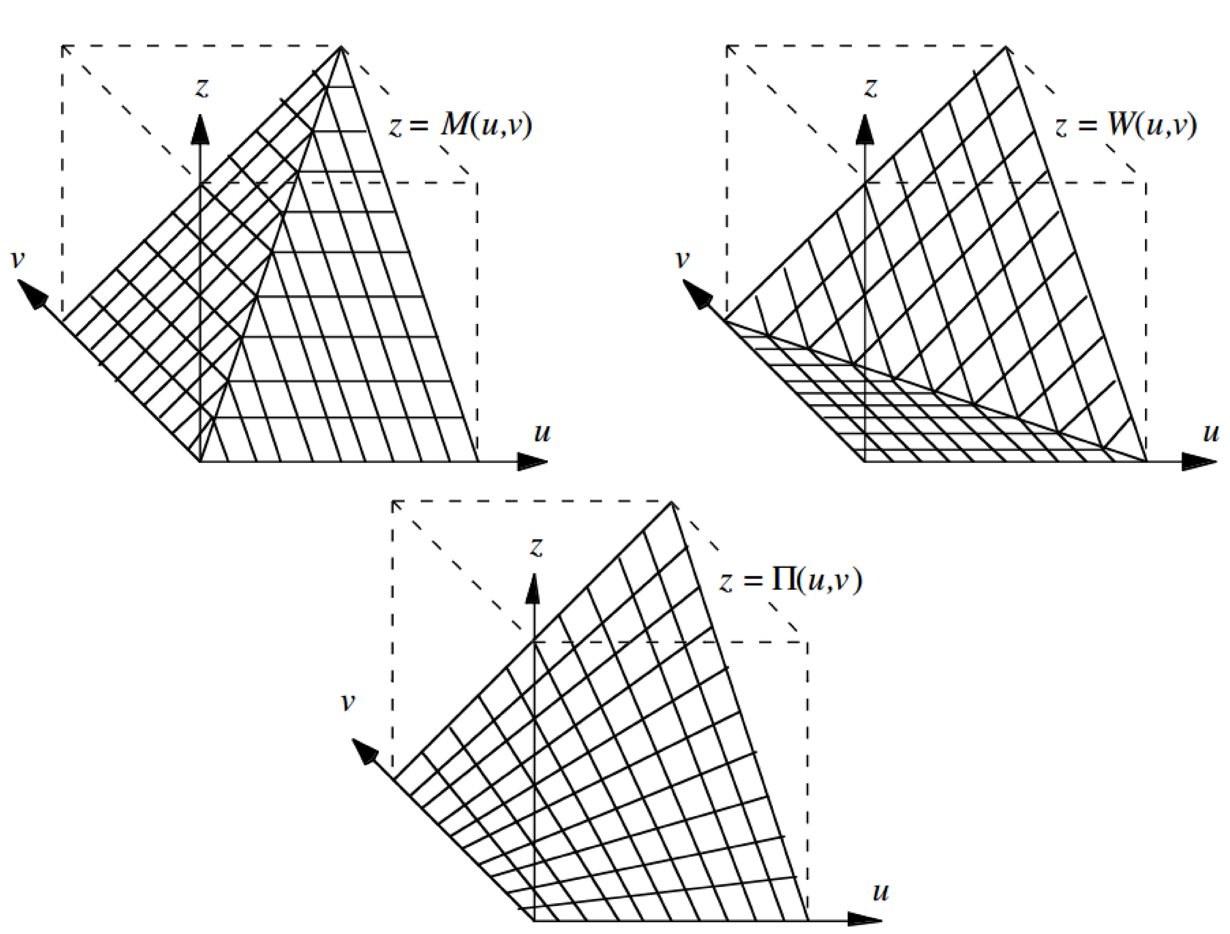

Figure 1: 3D and contour map of the copulas M, Π y W [Nelsen, 2006].

Bernstein copula

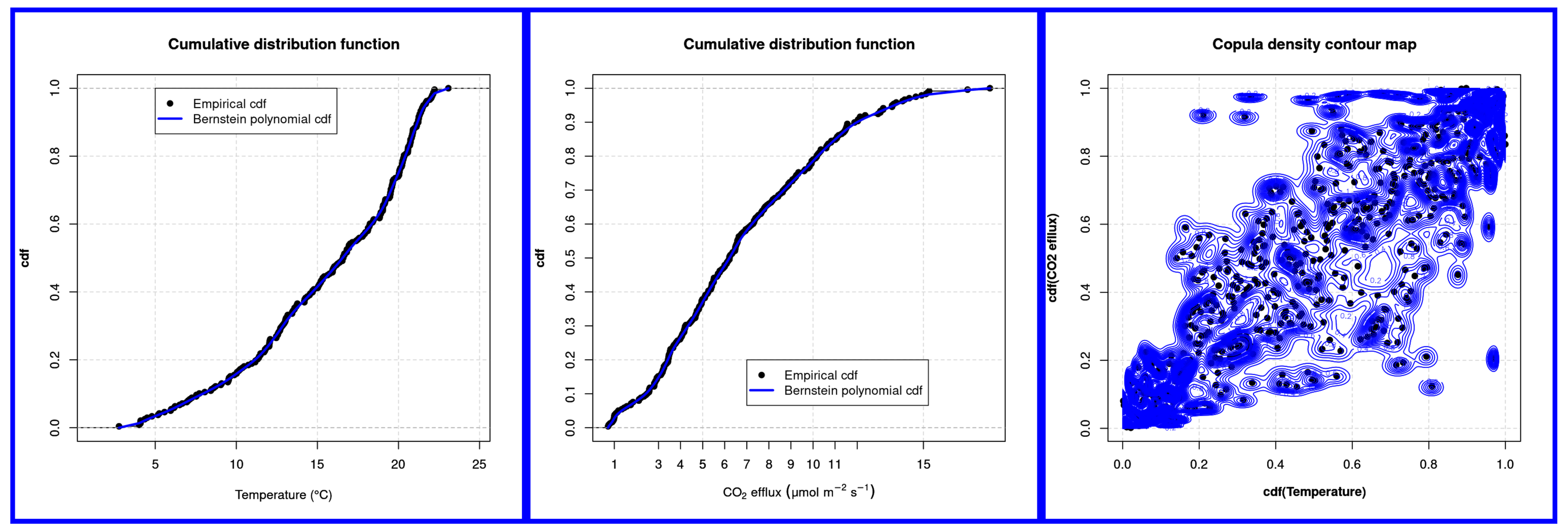

Figure 2: The univariate probability distribution function and the copula function are modeled using the Bernstein polynomial and the Bernstein copula.

Workflow

(1) Exploratory data analysis, (2) Copula-based dependency modeling, (3) Validation, (4) Application, (5) Uncertainty quantification.

Results

Univariate distribution

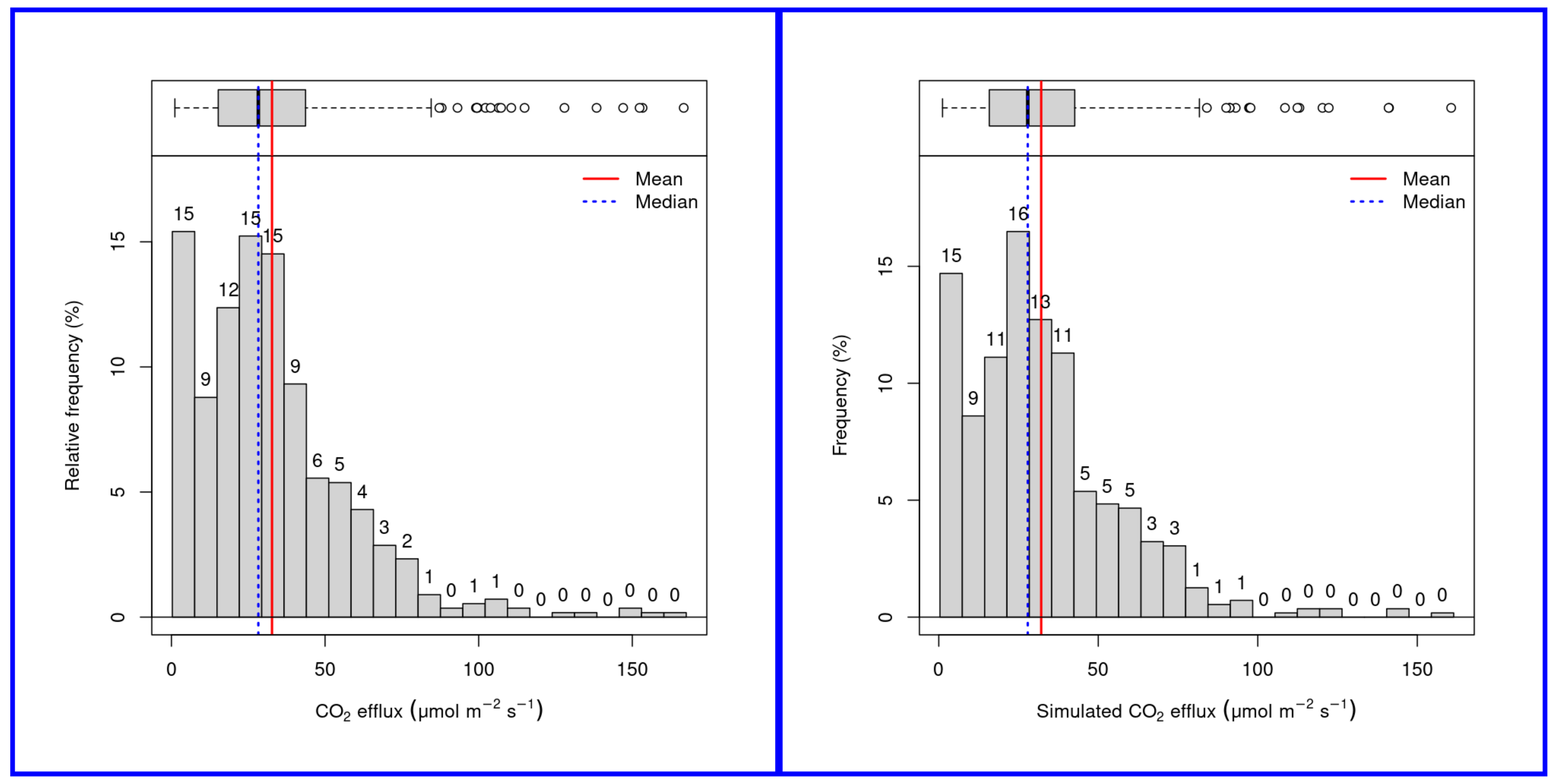

Figure 3: Histogram Boxplot: data vs conditional simulation.

Table 1: Statistics: data vs conditional simulation.

Statistics	CO_2 efflux ($\mu mol\ m^{-2}\ s^{-1}$)	Simulated CO_2 efflux ($\mu mol\ m^{-2}\ s^{-1}$)
Sample number	558	558
Minimum	1.1242	1.1755
1st. Quartile	15.2630	15.8817
Median	28.2857	27.8979
Mean	32.7306	32.1065
3rd. Quartile	43.6461	42.5038
Maximum	166.7340	160.6176
Rank	165.6098	159.4421
Interquartile Rank	28.3831	26.6221
Variance	626.3998	560.1549
Standard Deviation	25.0280	23.6676
Variation Coeff.	0.7647	0.7372
Skewness	1.6026	1.3984
Kurtosis	7.3086	6.3138

Bivariate distribution

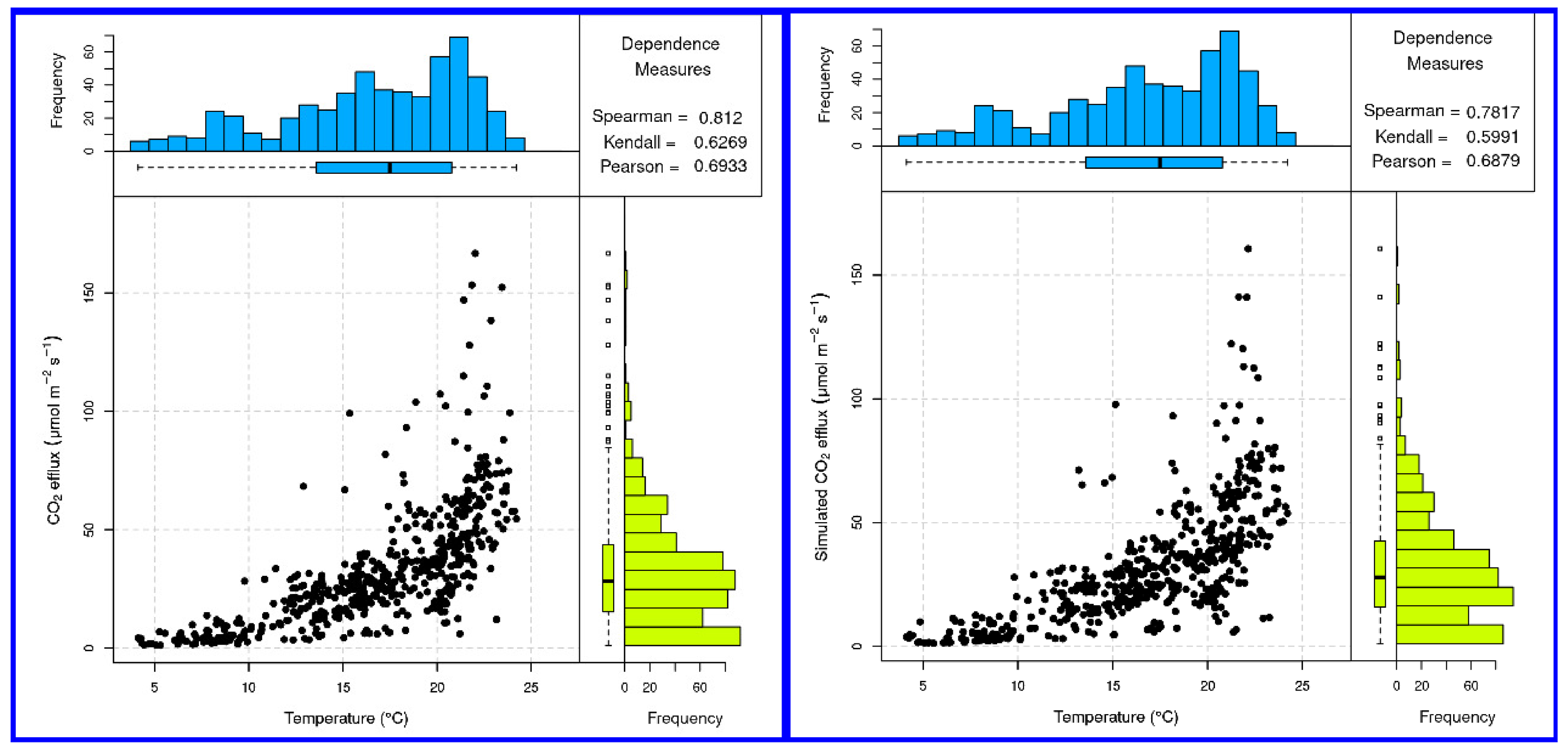

Figure 4: Scatterplot: data vs conditional simulation.

Temporal distribution

One hundred CO_2 efflux simulations conditioned to temperature shown in Figure 5; in which it allows informing in each temporal position t_i, the most likely value and its range of possible values, that is, the associated uncertainty range.

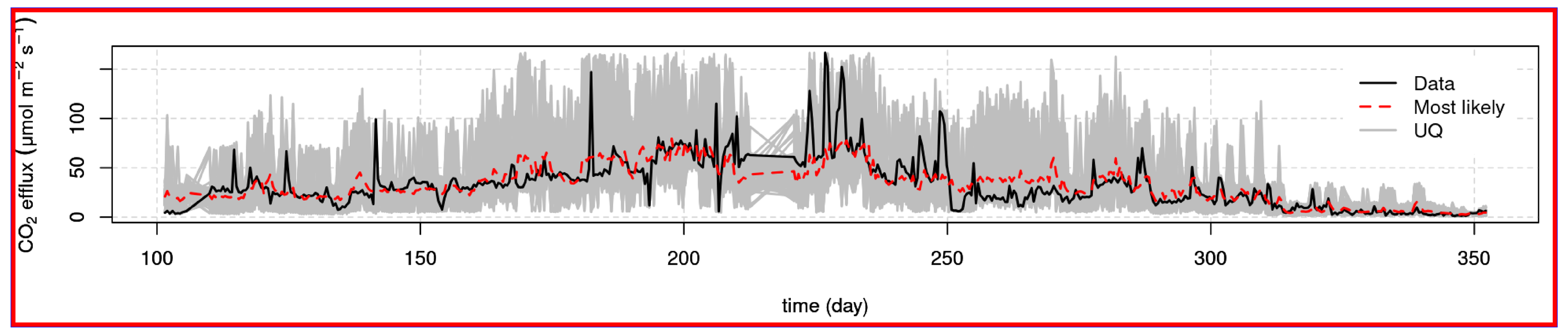

Figure 5: Temporal distribution: data vs conditional simulations.

Conclusions

Bernstein copula-based stochastic simulation method showed the ability to model, emit and reproduce behaviors such as univariate, bivariate and temporal of CO_2 efflux and temperature. In addition, this method allows to analyze the quantification of the uncertainty.

Extension and future work

1) Use of the Bernstein copula for subsampling optimization, 2) Extend to multivariate copulas ($n \geq 3$) using Vine copula, 3) Bayesian approaches to copula modeling.

References

[Le *et al.*, 2020] Le, V. H., Díaz-Viera, M. A., Vázquez-Ramírez, D., [del Valle-García], R., Erdely, A., y Grana, D. (2020). Bernstein copula-based spatial cosimulation for petrophysical property prediction conditioned to elastic attributes. *Journal of Petroleum Science and Engineering*, 193:107382.

[Nelsen, 2006] Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer-Verlag New York.

[Sklar, 1959] Sklar, A. (1959). Fonctions de répartition á n dimensions et leurs marges. *Publ. Inst. Statistique Univ. Paris*, 8:229–231.

Acknowledgment

This work was supported by a grant provided by NASA Carbon Monitoring System.