Modeling interactions within dairy farms using regular vine-copula models

Naomi Ouachene[1], Claudia Czado[2], Michael S. Corson[1], **Tristan Senga Kiessé**[1]

[1]UMR SAS, PHASE Division, INRAE

[2] School of Computation, Information and Technology & Munich Data Science Institute, Technical University of Munich, Germany

75th EAAP Annual Meeting, Florence (Italy), September 3, 2024

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

Introduction

Context

- 14.5-18% of global greenhouse gas (GHG) emissions come from **livestock production**, of which 75% are attributed to **ruminants** [5]
- **Interactions** among variables (e.g., milk production, GHG emissions, management practices, weather events) are **complex** to model [11]
- **Simulation models** exist (e.g., IFSM [13], DairyWise [14]), but they
  - need a large amount of input data [12] and pre-defined relations between variables [2] to provide output
  - are not useful for characterizing the type and shape of dependence structures that naturally exist among variables

Problem

How to **characterize** dependences (sometimes hidden or indirect) among descriptive variables of **several types** of farms ?

Goal

To provide a new way to describe **interactions** among a set of descriptive variables of farms by taking advantage of properties of the **copula method**.

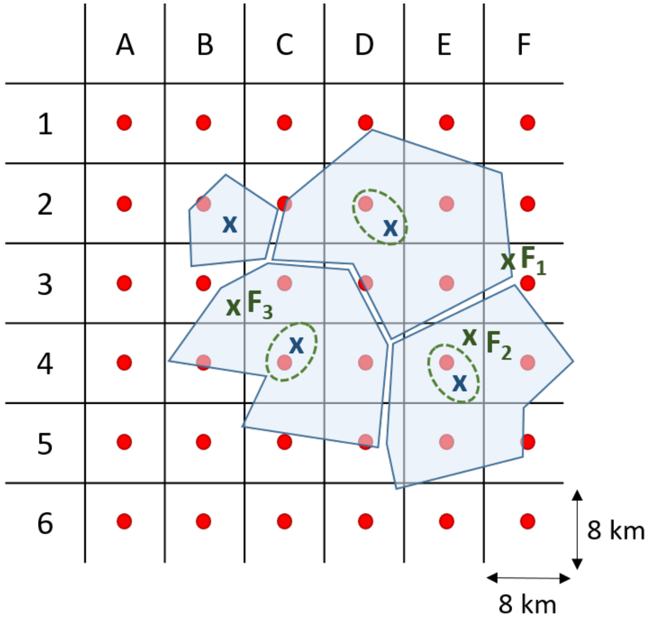
Data

- Data from the LIFE Carbon Dairy project of IDELE (French Livestock Institute) using the CAP'2ER method [7]
- **Raw** and **calculated** variables for 2347 dairy farms surveyed in 2013

- Response variables:
  - **Milk production**
  - **Greenhouse gas emissions (CO_2, CH_4 and N_2O)**

- Continuous explanatory variables, such as:
  - **Dry matter in the ration**
  - **Quantity of concentrated feed fed**
  - **Size of the farm**
  - **Number of animals**
  - **Meteorological indices**[1]

- Discrete explanatory variables, such as:
  - **Type of farm**
  - **Dominant cattle breed**
  - **Building type**



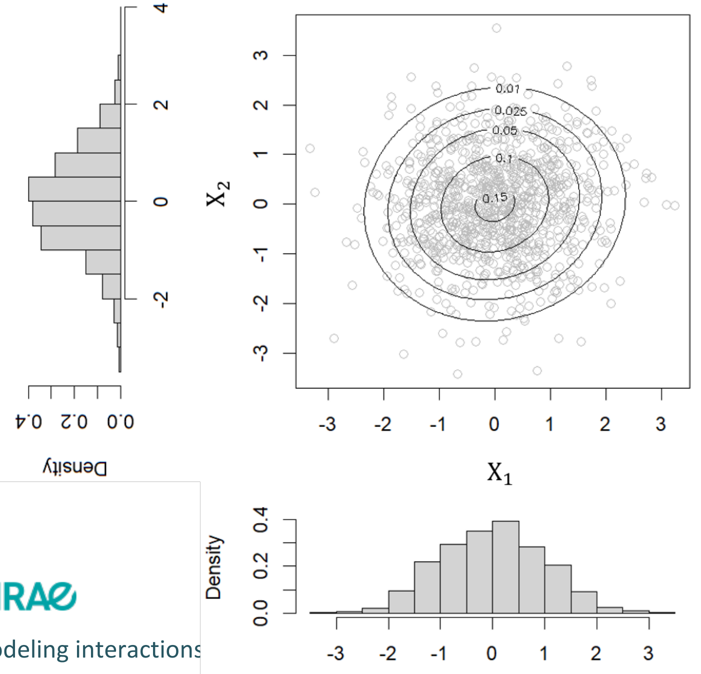
[1] The SAFRAN climate database, provided by Météo-France, was downloaded via the SICLIMA platform, developed by the AgroClim unit of INRAE

Vine copulas (1): definition of copulas

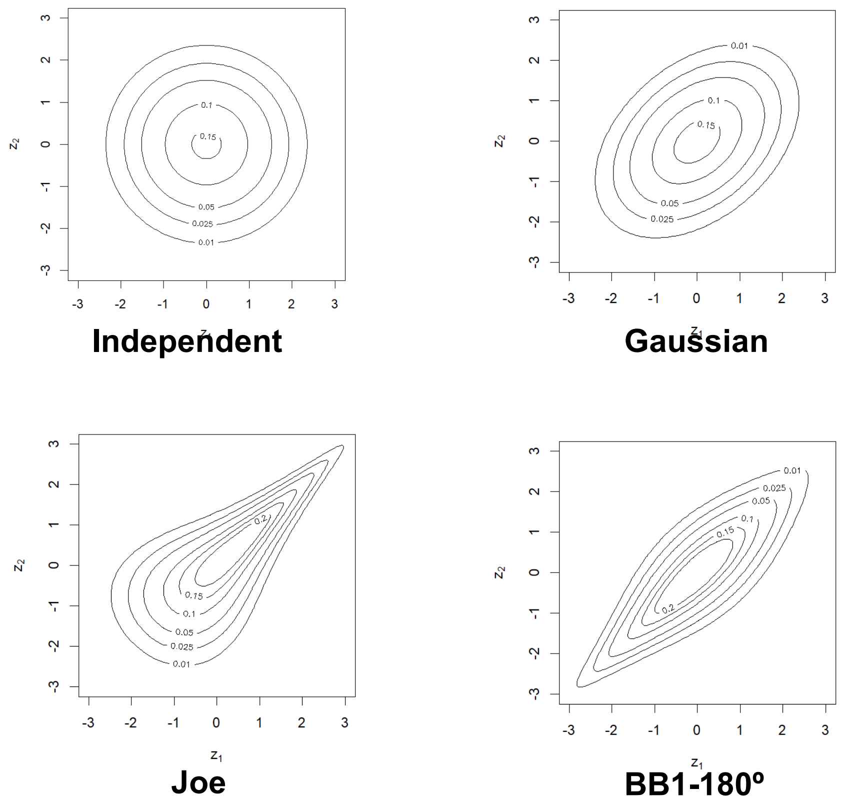
- Let $F_1, \cdots, F_d$ be cdfs of d continuous random variables $X_1, \cdots, X_d$ and F the joint cdf of $X = (X_1, \cdots, X_d)$
- According to Sklar's theorem [15], there exists a unique copula

  $\boldsymbol{C}: [0,1]^2 \rightarrow [0,1]$ such as $F(X_1, \cdots, X_d) = \boldsymbol{C}(F_1(x_1), \cdots, F_d(x_d))$

A copula model C formalizes the dependence structure between random variables, regardless of their distribution(s) [10, 6].



Contours of different types of bivariate copulas [8]



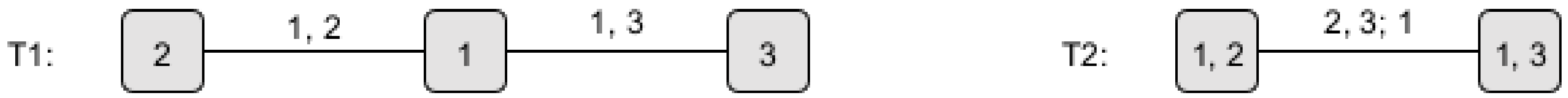
Vine copulas (2): pair-copula construction

Vines are a way to construct chains of bivariate copulas: the pair-copula construction (PCC) [1, 4]

Example for X_1, X_2, X_3

The joint density f of X_1, X_2, X_3 is defined such that [15, 4]

$$f = f_1 \times f_2 \times f_3 \times c_{12} \times c_{13} \times c_{23;1}$$



- f_1, f_2, f_3 are marginal densities of the 3 continuous random variables
- Vine copulas are chains of bivariate copulas that include **recursive conditioning** and allow for a **non-unique decomposition** [3]

Procedure for selecting the best R-vine structure by fitting the first tree to the strongest dependencies (*rvinecopulib* R package [9])

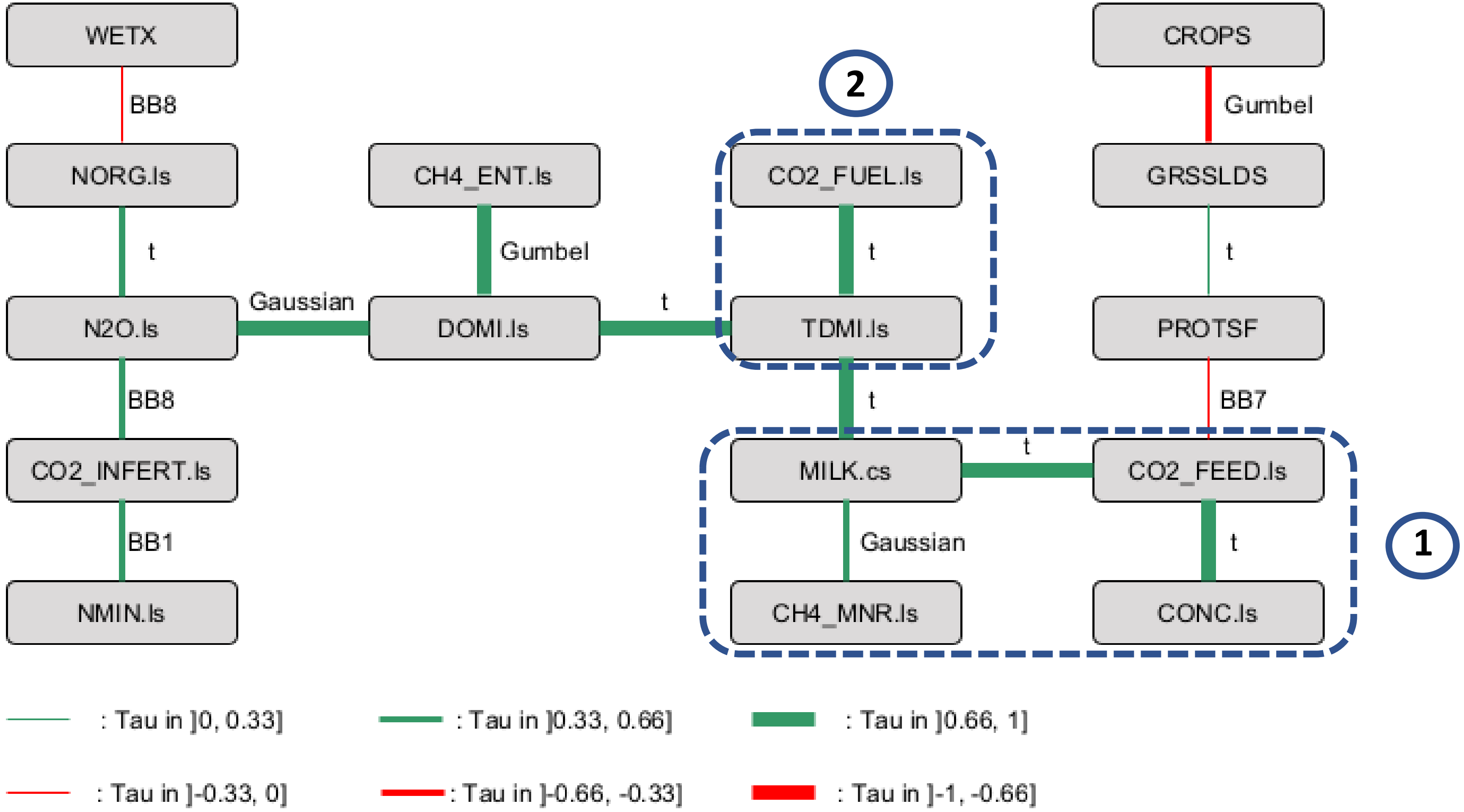
Application

Description of the variables analyzed for 2347 dairy farms surveyed in 2013. LU: livestock unit, UAA: utilized agricultural area (ha), CH_4: methane, CO_2: carbon dioxide

Variable	Description (unit)	Standardized by
MILK.cs	Milk production ($l.cow^{-1}$)	UAA
CH4_ENT.ls	Enteric CH_4 emissions (kg CO_2-eq.LU^{-1})	UAA
CH4_MNR.ls	CH_4 emissions due to manure management (kg CO_2-eq.LU^{-1})	UAA
CO2_FEED.ls	CO_2 emissions due to the purchase of feed (kg CO_2-eq.LU^{-1})	UAA
CO2_FUEL.ls	CO_2 emissions due to on-farm fuel consumption (kg CO_2-eq.LU^{-1})	UAA
CO2_INFERT.ls	CO_2 emissions due to the purchase of inorganic fertilizers (kg CO_2-eq.LU^{-1})	UAA
N2O.ls	Total nitrous oxide emissions (kg CO_2-eq.LU^{-1})	UAA
CROPS	Area of crops (proportion of the UAA)	-
GRSSLDS	Area of grasslands (proportion of the UAA)	-
NMIN.ls	Inorganic nitrogen fertilizers applied ($kg.ha^{-1}$)	LU
NORG.ls	Organic nitrogen fertilizers applied ($kg.ha^{-1}$)	LU
TDMI.ls	Total dry matter ingested ($kg.LU^{-1}$)	UAA
DOMI.ls	Digestible organic matter in the ration ($kg.LU^{-1}$)	UAA
CONC.ls	Concentrated feed fed ($kg.LU^{-1}$)	UAA
PROTSF	Protein self-sufficiency (proportion)	-
WETX	Soil wetness index of the area (proportion)	-

Results (1)

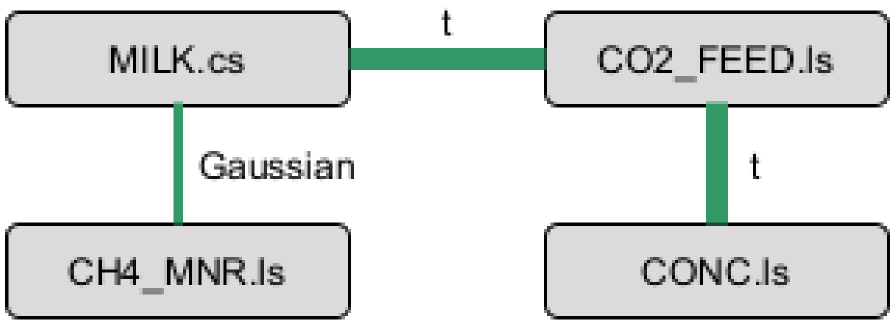
First tree of the R-vine (all farms included)



Results (2)

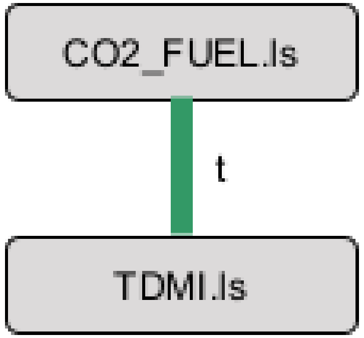
Dependences between:

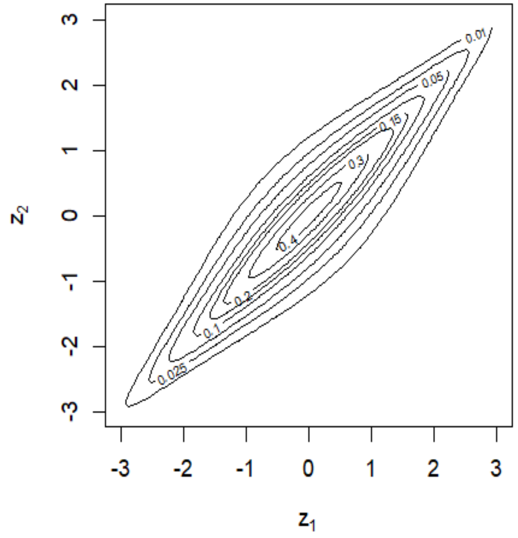
- CO_2 emissions due to the purchase of feed [CO2_FEED.ls] and the amount of concentrated feed fed to animals [CONC.ls]:
  - a large proportion of purchased feed would be concentrated feed

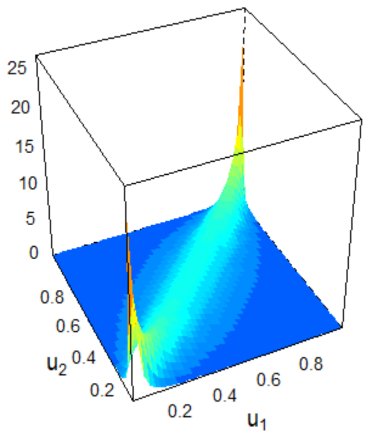


- milk production [MILK.cs] and CH_4 emissions associated with manure management [CH4_MNR.ls]:
  - strategies to produce milk influence herd management, time spent inside the barn and the type of manure management

Results (3)







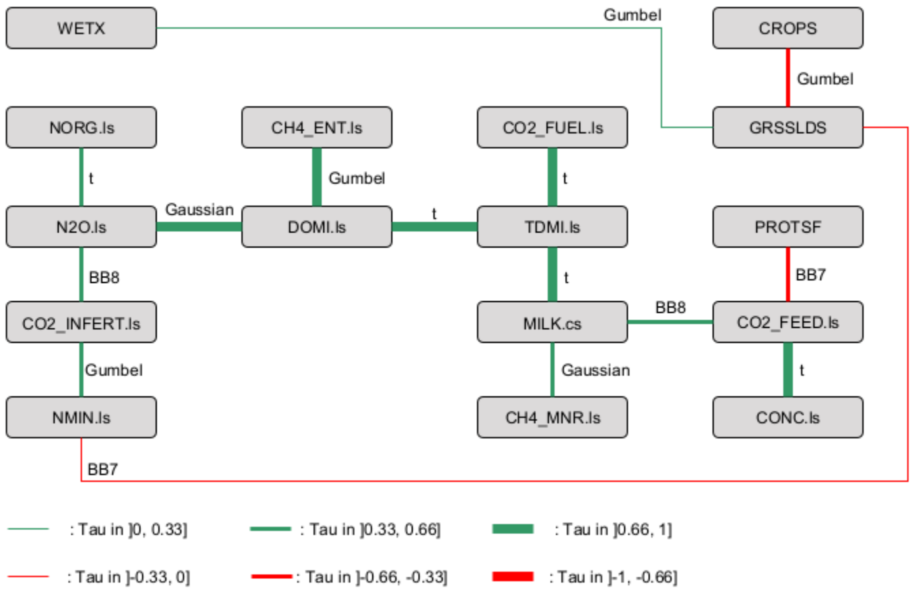
Direct dependence between the amount of dry matter ingested [TDMI.ls] and the CO_2 emissions due to fuel consumption [CO2_FUEL.ls]:

- the amount of dry matter ingested comes from the forage yield, which requires a certain amount of fuel consumption

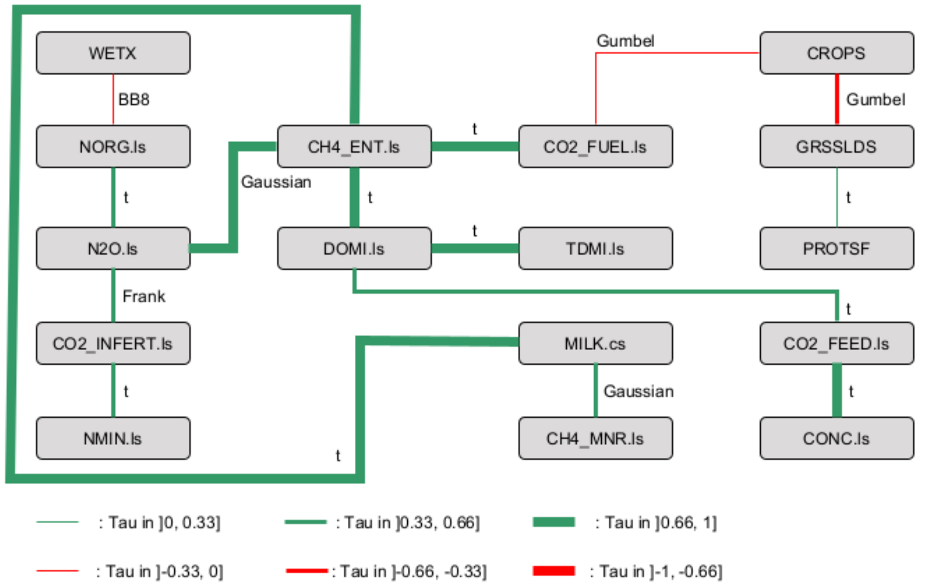
Results (4)

First tree of the R-vine

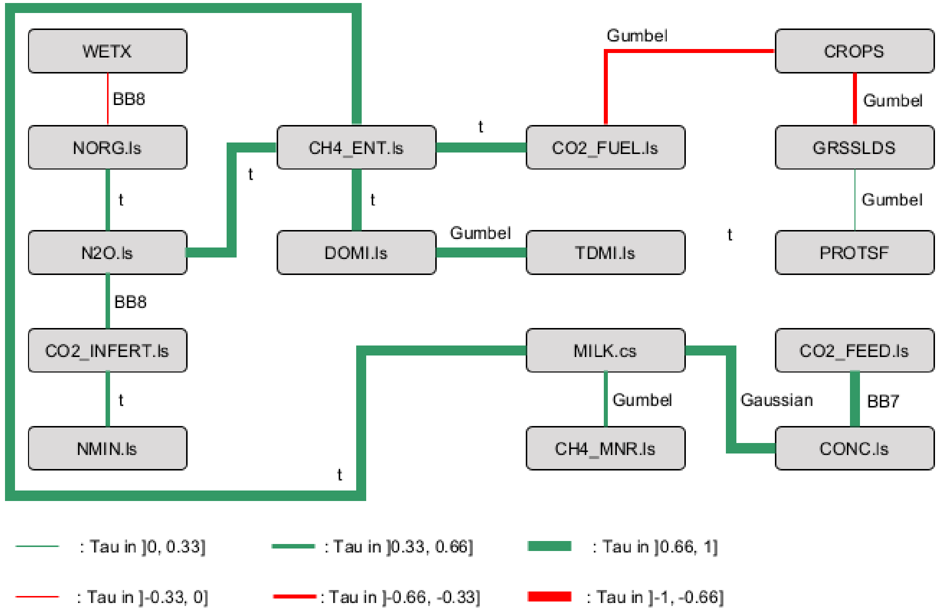
(farms grouped by milk production per cow)



A: milk production ≤ 7500 kg per cow (799 farms)



B: milk production of 7500-8500 kg per cow (819 farms)

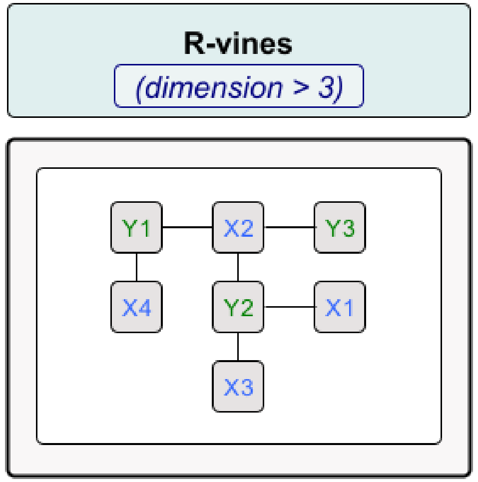


C: milk production > 8500 kg per cow (729 farms)

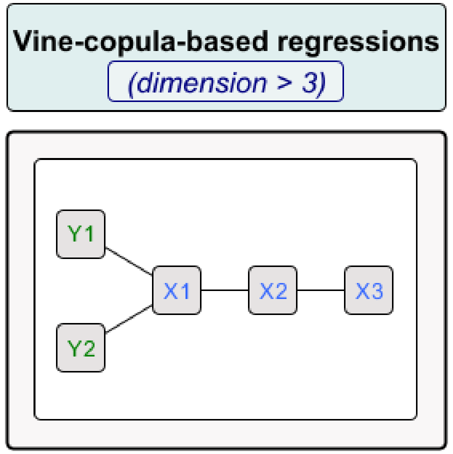
- Differences in dependence structures due to differences in farm productivity

Conclusion & Perspectives

Conclusion



- **Map** and **characterize** multiple dependences by taking advantage of **bivariate-copula** properties
- Capture the behavior of **non-conventional** farms using extreme-value copulas
- Assessments of **all farms** and **by type of farm**, which showed differences as a function of management strategies



Perspectives

- Identify variables X_i that explain response variables Y_i the most

Thank you for your attention

References

[1] Kjersti Aas et al. "Pair-copula constructions of multiple dependence". In: Insurance: Mathematics and Economics 44.2 (2009), pp. 182–198. ISSN: 0167-6687.

[2] Xavier Chardon. "Evaluation environnementale des exploitations laitières par modélisation dynamique de leur fonctionnement et des flux de matière : Développement et application du simulateur MELODIE". Ph.D. thesis. AgroParisTech (2008).

[3] Claudia Czado. Analyzing Dependent Data with Vine Copulas. 1st ed. Springer Cham (2019).

[4] Claudia Czado and Thomas Nagler. "Vine copula based modeling". In: Annual Review of Statistics and Its Application 9 (2022), pp. 453–477.

[5] Bertrand Dumont et al. "Livestock farming in Europe: a diversity of services and impacts". In: INRA Productions Animales 30.4, SI (2017), pp. 271–272. ISSN: 0990-0632.

[6] Christian Genest and Anne-Catherine Favre. "Everything you always wanted to know about copula modeling but were afraid to ask". In: Journal of Hydrologic Engineering 12.4 (2007), pp. 347–368.

[7] IDELE. Calcul Automatisé des Performances Environnementales pour des Exploitations Responsables. URL: https://cap2er.fr/Cap2er/.

[8] Saralees Nadarajah, Emmanuel Afuecheta, and Stephen Chan. "A Compendium of Copulas". In: Statistica 77.4 (2017), pp. 279–328.

[9] Thomas Nagler and Thibault Vatter. rvinecopulib: high performance algorithms for vine copula modeling. R package version 0.6.3.1.1. 2020. URL: https://CRAN.R-project.org/package=rvinecopulib.

References

[10] Roger Nelsen B. 2nd ed. An Introduction to Copulas. Springer Series in Statistics (2006).

[11] Sylvain Pellerin et al. “Identifying cost-competitive greenhouse gas mitigation potential of French agriculture”. In: Environmental Science & Policy 77 (2017), pp. 130–139. ISSN: 1462-9011.

[12] C. Rosenzweig et al. “The Agricultural Model Intercomparison and Improvement Project (AgMIP): Protocols and pilot studies”. In: Agricultural and Forest Meteorology 170 (2013), pp. 166–182. ISSN: 0168-1923.

[13] Clarence Alan Rotz et al. The Integrated Farm System Model. USDA ARS Washington, DC (2022).

[14] RLM Schils et al. “DairyWise, a whole-farm dairy model”. In: Journal of Dairy Science 90.11 (2007), pp. 5334–5346. ISSN: 0022-0302.

[15] Abe Sklar. “Fonctions de répartition à n dimensions et leurs marges”. In: Publication de l’Institut de Statistique de L’Université de Paris 8 (1959), pp. 229–231.