Using vine copulas to explore relationships between variables determining deforestation in the Brazilian Amazon

GIScience for environmental change - use cases in the Brazilian Amazon
15.06.2011

Benedikt Gräler
Institute for Geoinformatics
University of Münster
http://ifgi.uni-muenster.de/graeler

Deforestation

- the process of deforestation is complex
- the size of the area is hard to grasp
- the number of potential explanatory variables is large

Probabilistic speaking:
We need to capture/model/understand a multi-variate random process over a large spatial/spatio-temporal region.

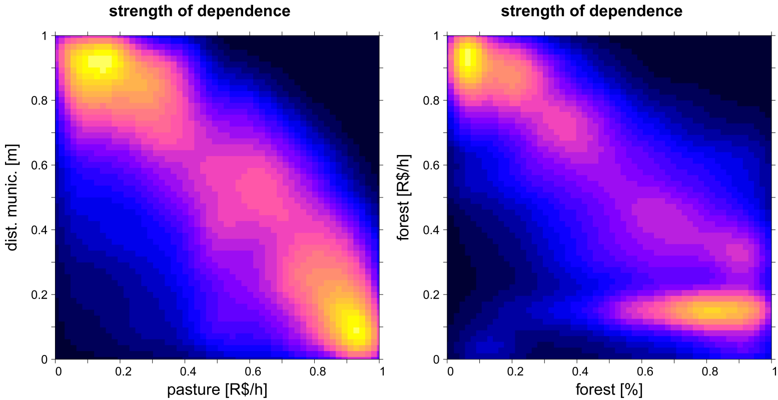
ifgi
Institute for Geoinformatics
University of Münster

Additional challenges

- marginal distributions are of various kinds (i.e. non-Gaussian)
- dependence structures are asymmetric (i.e. non-elliptical)
- some variables exhibit joint tail-dependence

Thus, the assumption of an underlying Gaussian process is questionable.

smoothed scatter plots showing strength of dependece



ifgi
Institute for Geoinformatics
University of Münster

Vine Copulas

- the theory of *vine copulas*[1] allows to break up a multi-variate distribution into a set of *bivariate* ones
- any bivariate one can be chosen freely
- margins and dependence structures (represented by *copulas*) can be handled independently

[1] see [1] for further details

C-vines

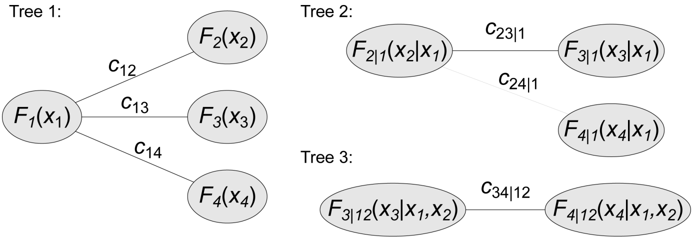


Figure: a 4-dimensional canonical vine

$$f_C(x_1,\dots,x_4) = \prod_{k=1}^{4} f(x_k) \prod_{j=1}^{4-1} \prod_{i=1}^{4-j} c_{j,j+i|1\dots j-1}(\cdots)$$

D-vines

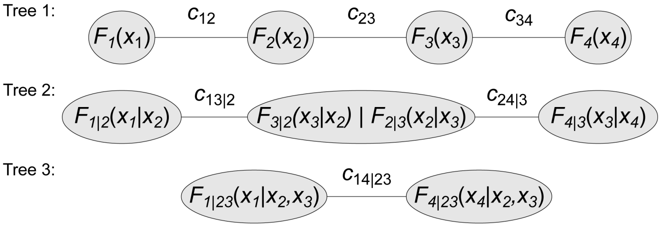


Figure: a 4-dimensional D-vine

$$f_D(x_1,\dots,x_4) = \prod_{k=1}^{4} f(x_k) \prod_{j=1}^{4-1} \prod_{i=1}^{4-j} c_{i,i+j|i+1\dots i+j-1}(\cdots)$$

R-vines

- lie "in between" C and D-vines
- are by far more flexible

One has to choose

- there are many different ways to decompose a multivariate distribution into bivariate ones
- typically, different choices will give different results

ifgi
Institute for Geoinformatics
University of Münster

Maximum Spanning Trees

Definition

Given a connected, undirected and weighted graph. A *maximum spanning tree* is a cycle-free subgraph that connects all vertices and its sum over all edges is larger or equal than for any other such cycle-free subgraph.

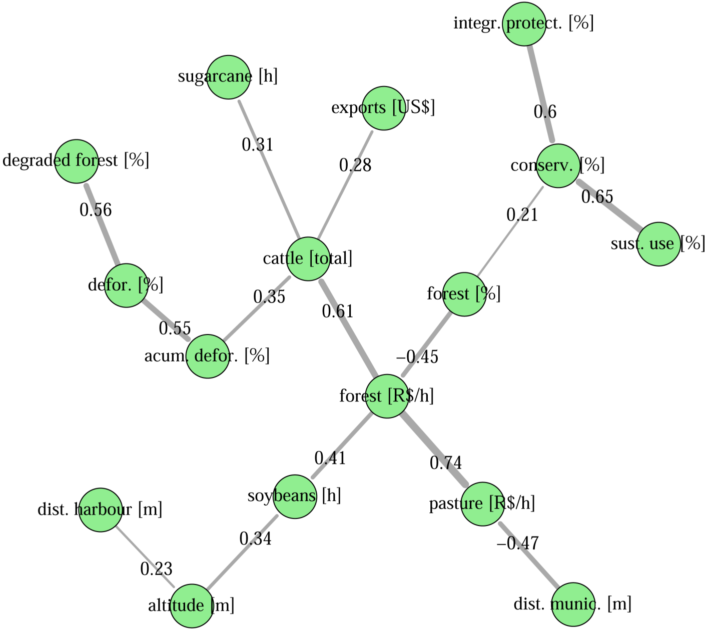
Maximum spanning trees and R-vines

We derive a maximum spanning tree by (following [2]):

1. calculating Kendall's tau for all pairs of variables of interest
2. building a fully connected and undirected graph over all these variables
3. assigning Kendall's tau as weights for the corresponding edges
4. calculating the maximum spanning tree for this graph

ifgi
Institute for Geoinformatics
University of Münster

A maximum spanning tree over deforestation variables



References

- Aas, Kjersti, Claudia Czado, Arnoldo Frigessi & Henrik Bakken (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics, 44*, 182 - 198.
- Dißmann, J., E. C. Brechmann, C. Czado, and D. Kurowicka (2011). Selecting and estimating regular vine copulae and application to financial returns. *Submitted for publication.*

ifgi
Institute for Geoinformatics
University of Münster