Chapter 5
Amazon copula

Seminar *Spatio-temporal dependence*,
07.02.2011 - 11.02.2011

Benedikt Gräler
Institute for Geoinformatics
University of Muenster

Outline

1. **a few strange copulas**
   - asymmetric copulas
   - copulas for zero inflated data

ifgi
Institute for Geoinformatics
University of Münster

Outline

ifgi
Institute for Geoinformatics
University of Münster

Outline

ifgi
Institute for Geoinformatics
University of Münster

An asymmetric copula - ASC I

ifgi
Institute for Geoinformatics
University of Münster

So far, all the families we looked at were symmetric copula families. A two parameter asymmetric copula is given as follows:

Example

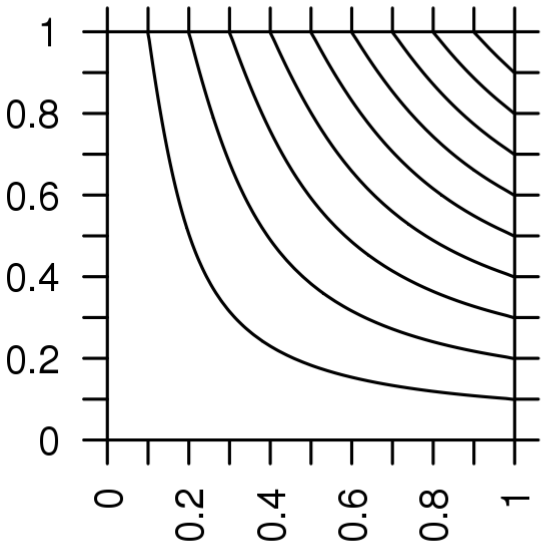
A family of copulas which is not symmetric is the following

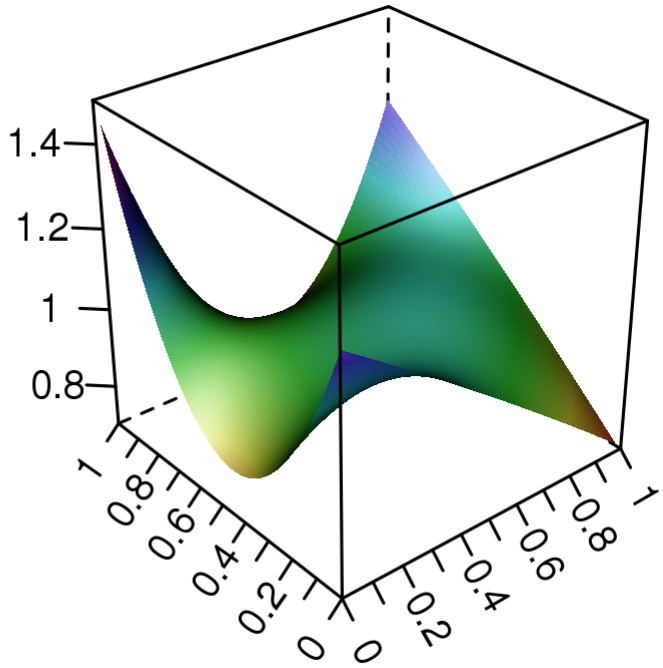
$$C^A_{ab}(u,v) = uv + uv(1-u)(1-v)\big((a-b)v(1-u)+b\big)$$

for all $|b| \leq 1$ and $\big(b-3-\sqrt{9+6b-3b^2}\big)/2 \leq a \leq 1$ with $a \neq b$ (see Example 3.16 in [Nelsen 2006]). We will denote this two parameter family as family of asymmetric copulas (ASC).

An asymmetric copula - ASC II

Contourplot and density for one pair of parameters are illustrated below:





An asymmetric copula - ASC III

ifgi
Institute for Geoinformatics
University of Münster

An asymmetric copula can be estimated using a combination of the inversion of Kendall's tau/Spearman's rho and a maximum likelihood estimation [Gräler 2009]. Where the parameter a can be estimated by:

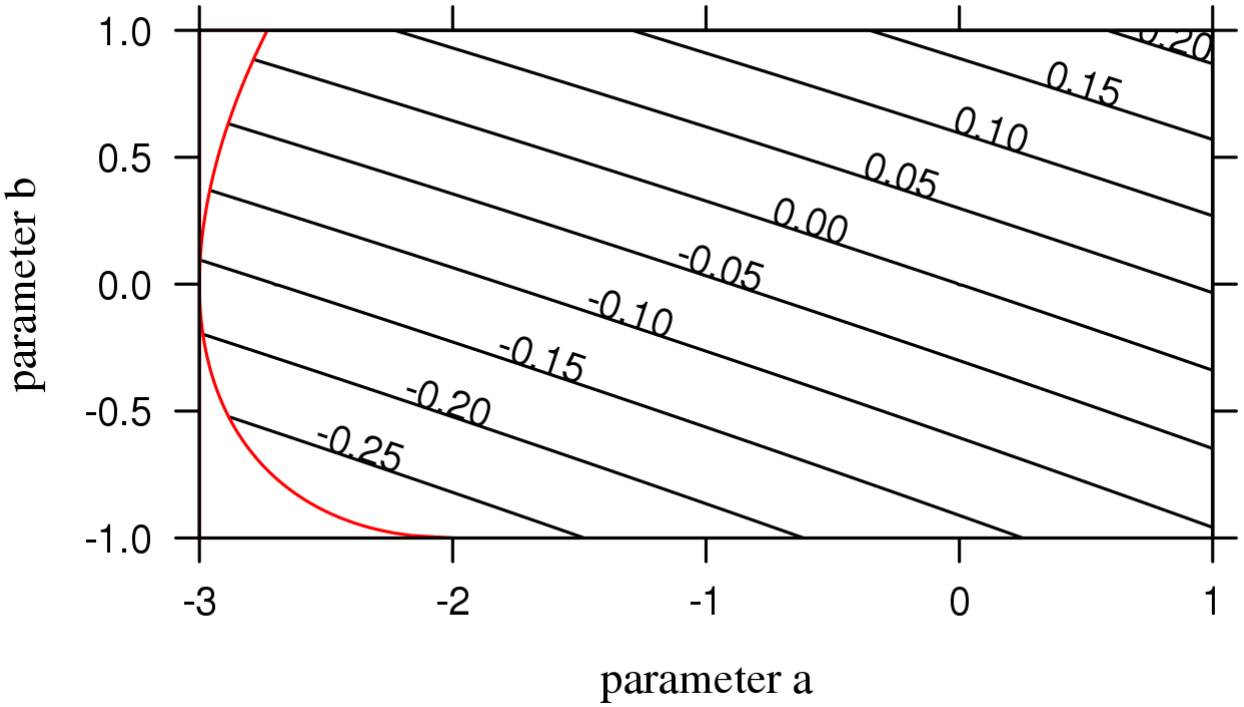
$$\hat{a} = \frac{450\tau - 75b + b^2}{25 - b}$$

or

$$\hat{a} = 12\rho - 3b$$

The parameter b than needs to be fitted using a maximum likelihood estimator. Valid parameters can only be obtained within the following region:

An asymmetric copula - ASC IV



Where the contour lines denote the value of Kendall's tau.

Zero inflated data I

ifgi
Institute for Geoinformatics
University of Münster

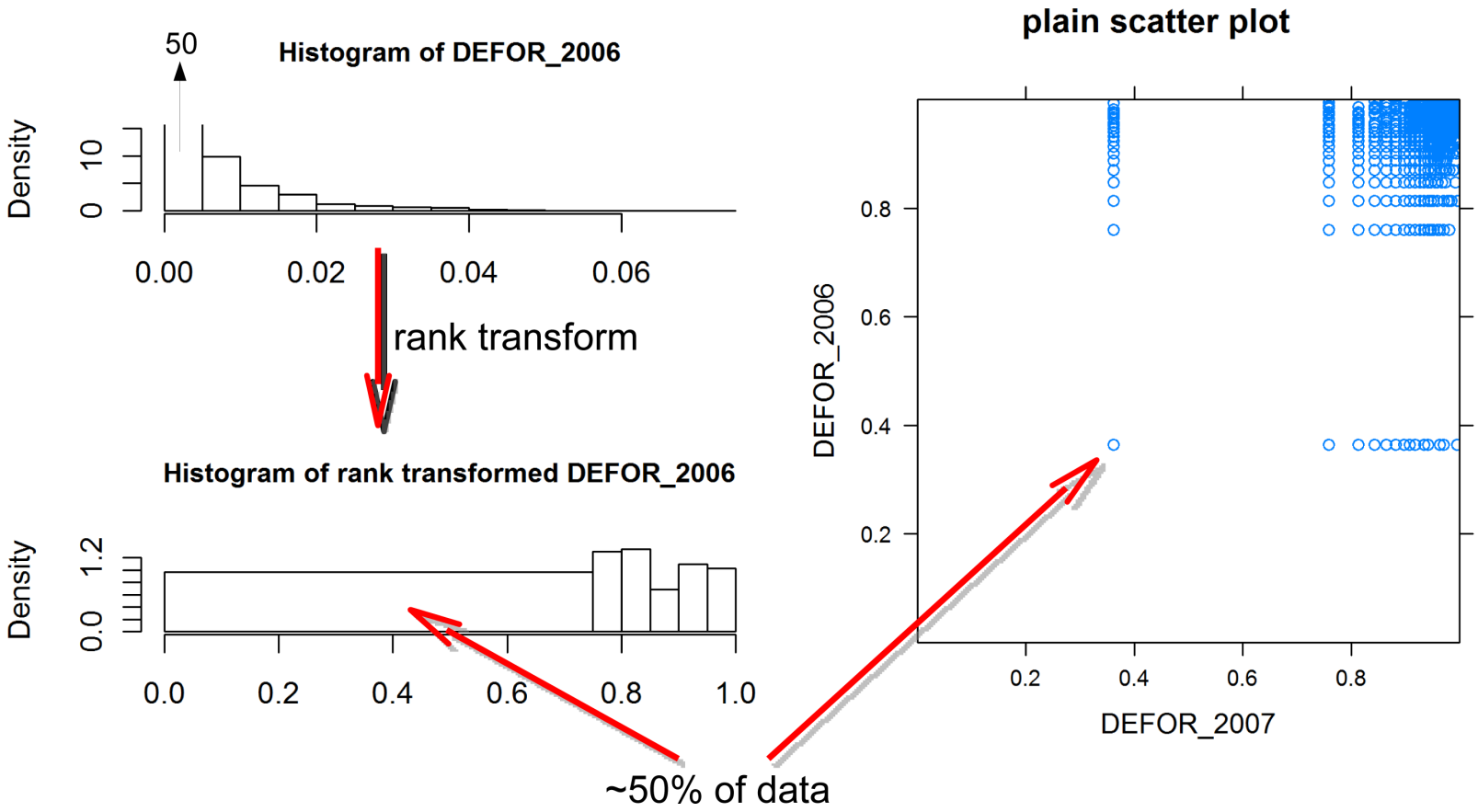
In several applications one will find a huge amount of 0s (or very small values) in a sample. This is the case for example for

- rainfall data
- nuclear radiation
- deforestation

This leads to scatter plots where a large quantity of observations is concentrated in a single point or line.

But, copulas assume continuous, equally spread data instead.

Zero inflated data II



ifgi
Institute for Geoinformatics
University of Münster

copulas for zero inflated data - TMC I

An approach to solve this problem is by introducing *truly mixed copulas* (*TMC*) [Gräler et al. 2010].

The unit square is broken up into four areas: the lower left rectangle denoting the zero-zero pairs, the top left and lower right rectangles denoting the zero-non-zero and non-zero-zero pairs and the top right corner which can be rescaled and modeled as copula.

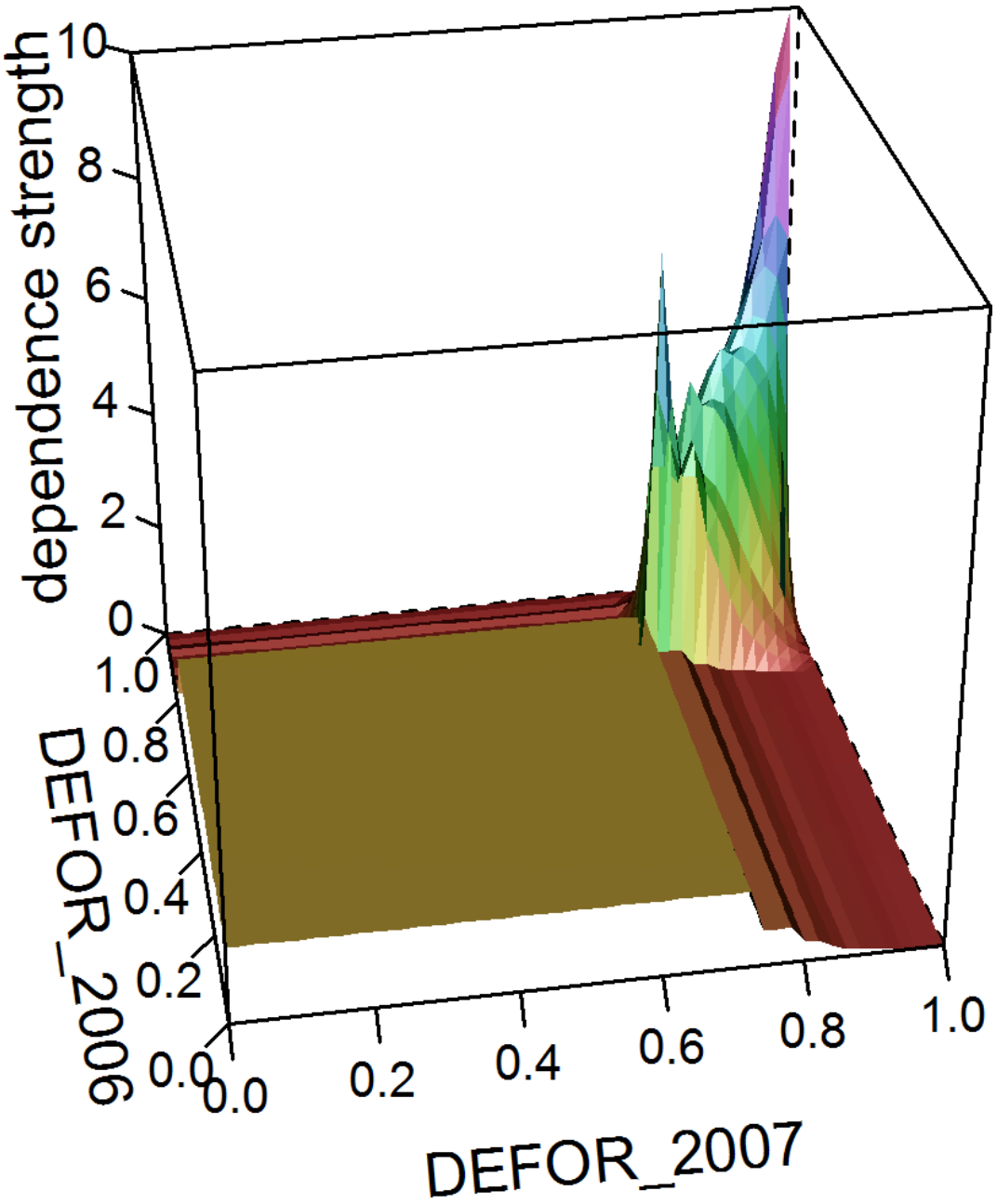
copulas for zero inflated data - TMC II

The rescaling is done in a way that the joint bivariate function is a copula again maintaining the mass relations and copula properties:

To achieve this, we need to estimate inner marginal functions and counter parts such that both add up to a constant 1.

A truly mixed copula density might look like:

copulas for zero inflated data - TMC III



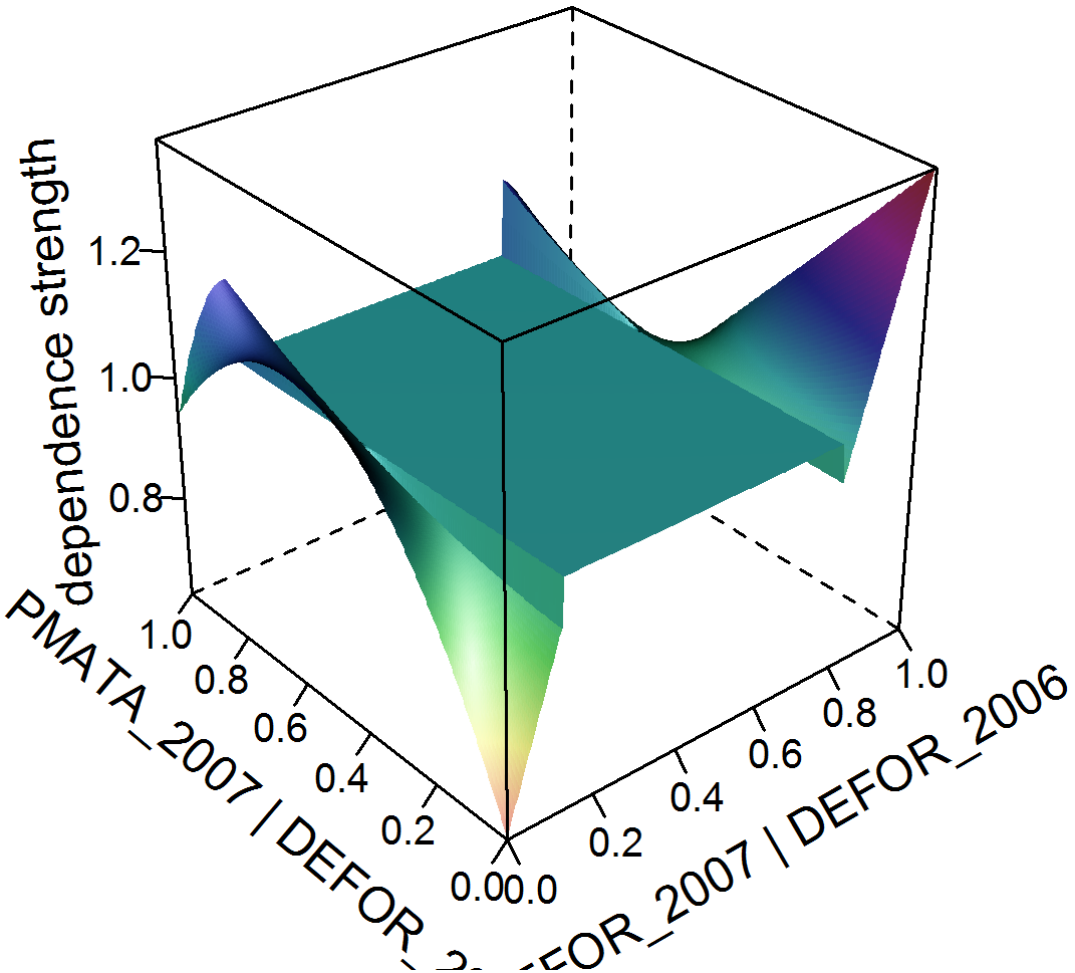
cutted copulas I

Instead of in one corner one might find a big bunch of values some where in the middle of the unit interval.

This part can be cut out according to its mass and inserted after the estimation process [Gräler et al. 2010].

Depending on the distribution of this cut-out, a distribution function might be necessary. The cut copula looks like

cutted copulas II



Benedikt Gräler

ifgi
Institute for Geoinformatics
University of Münster

Deforestation of the Amazon

The following study is published in [Gräler et al. 2010] and was presented at the Research Symposium GIScience for Environmental Change, November 27, 2010, Campos do Jordão (São Paulo), Brazil.

ifgi
Institute for Geoinformatics
University of Münster

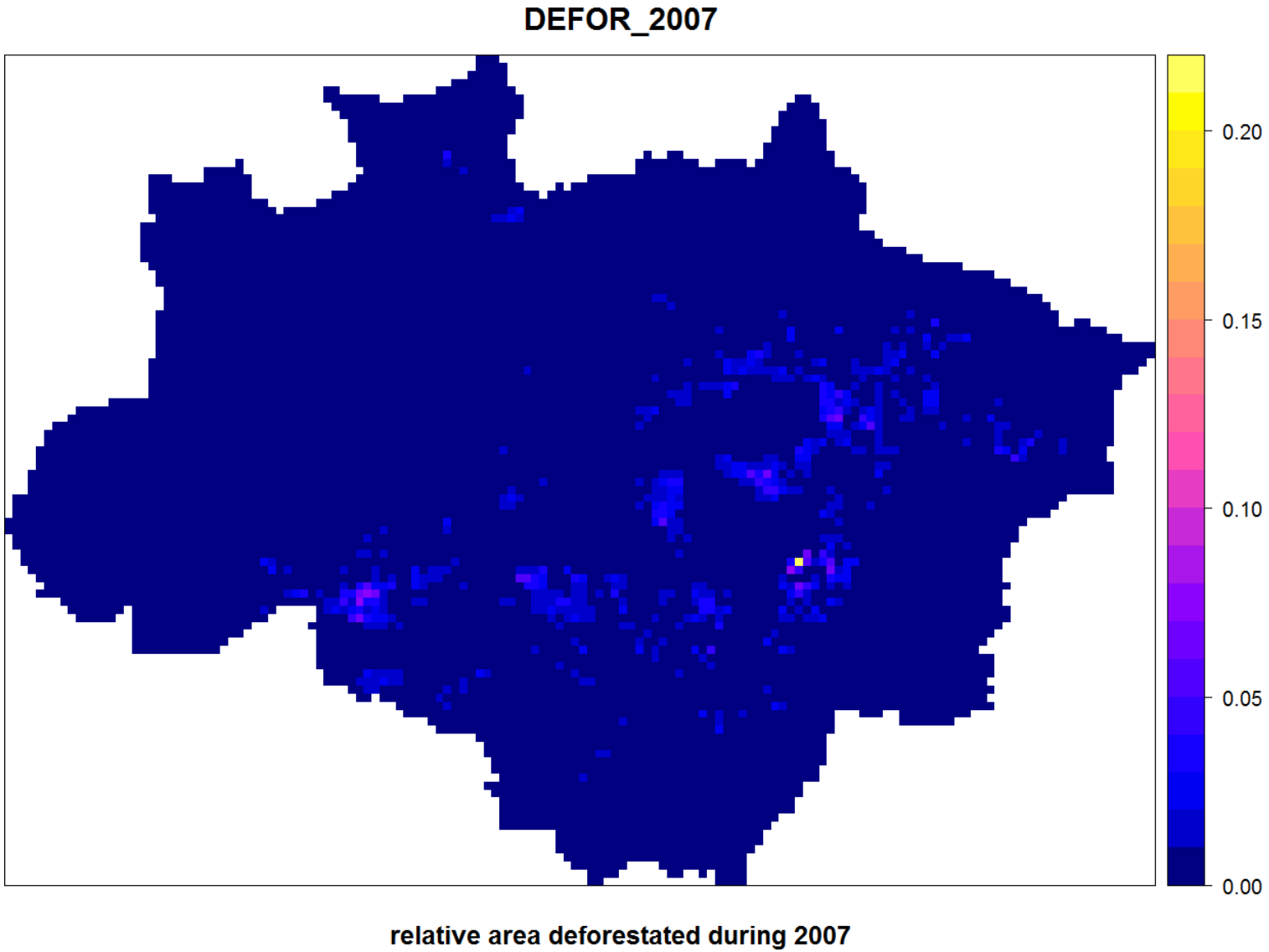
Deforestation in the Amazon

Benedikt Gräler

ifgi
Institute for Geoinformatics
University of Münster

The data I

The amount of yearly deforested area per raster cell is calculated by INPE, Brazil.



The data II

Additional variables present are e.g.

- demographic information
- altitude
- preserved ares
- price of forest land
- area of sugarcane or soy beans

and many more!

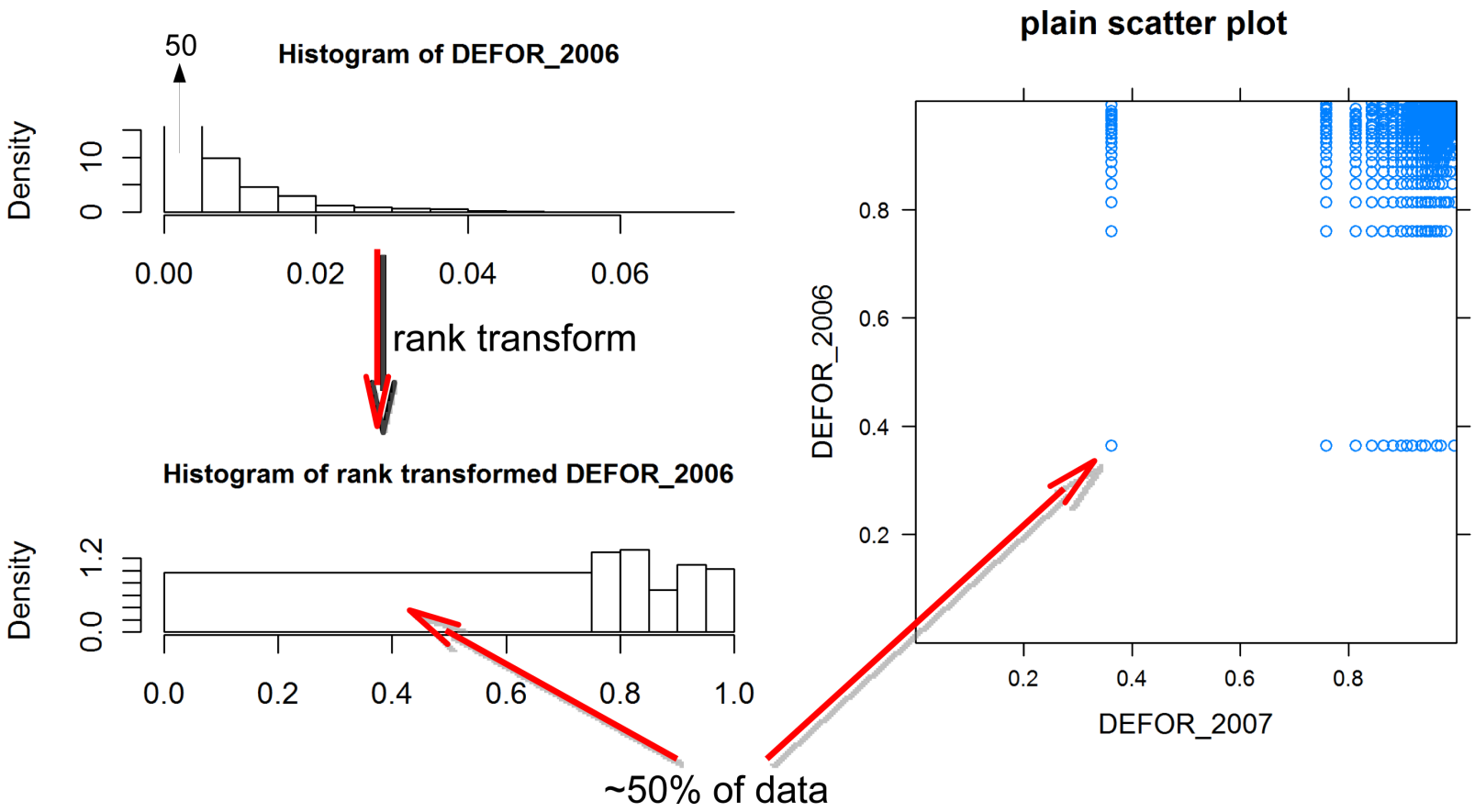
We will investigate the three dimensional random process given by:

$$\text{defores. } 2006 \approx \text{defores. } 2007 \approx \text{price of forest } 2007$$

zero inflated data

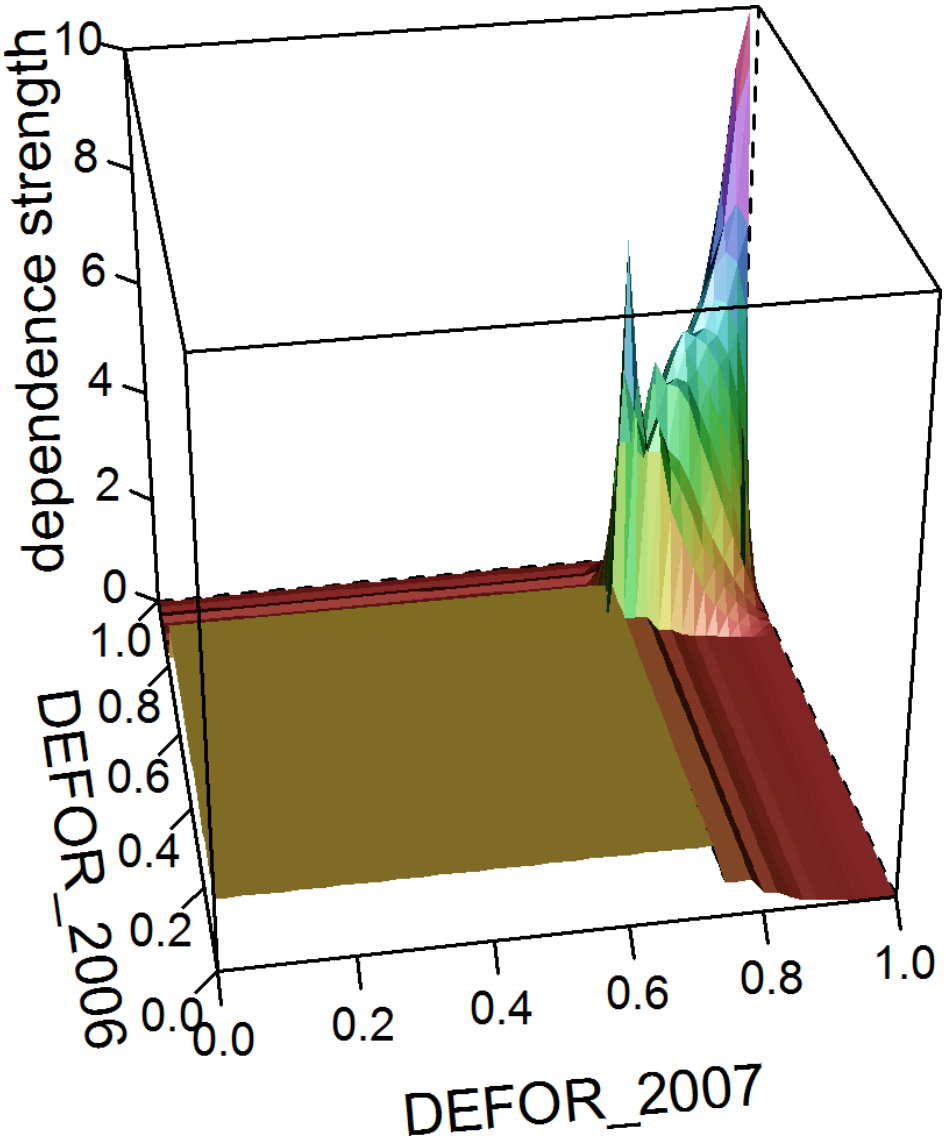
ifgi
Institute for Geoinformatics
University of Münster

About 50% of the data is 0:



truly mixed copulas

In order to cope with this massive amounts of 0s we cut the copula in four parts: both margins are 0, either one is 0 and none of both is 0.

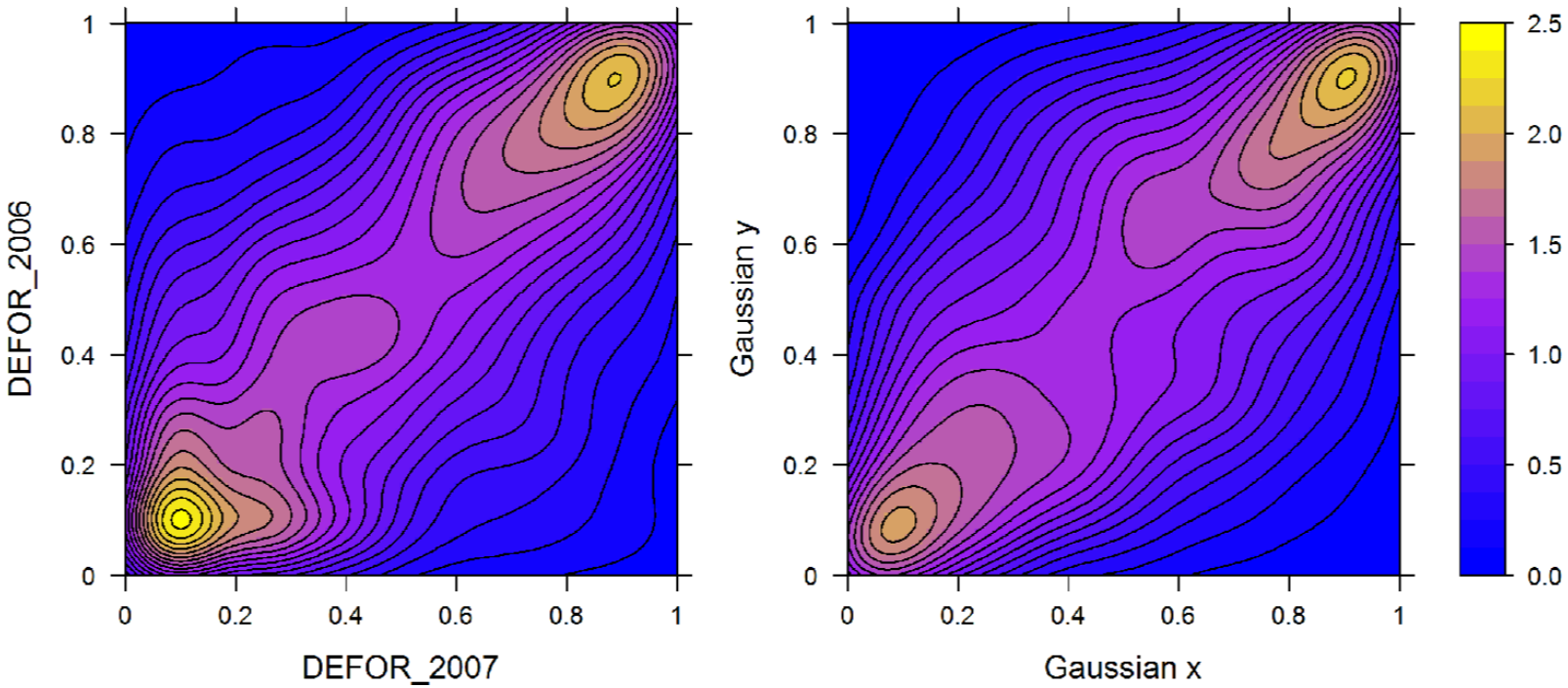


ifgi
Institute for Geoinformatics
University of Münster

fitting the TMC

ifgi
Institute for Geoinformatics
University of Münster

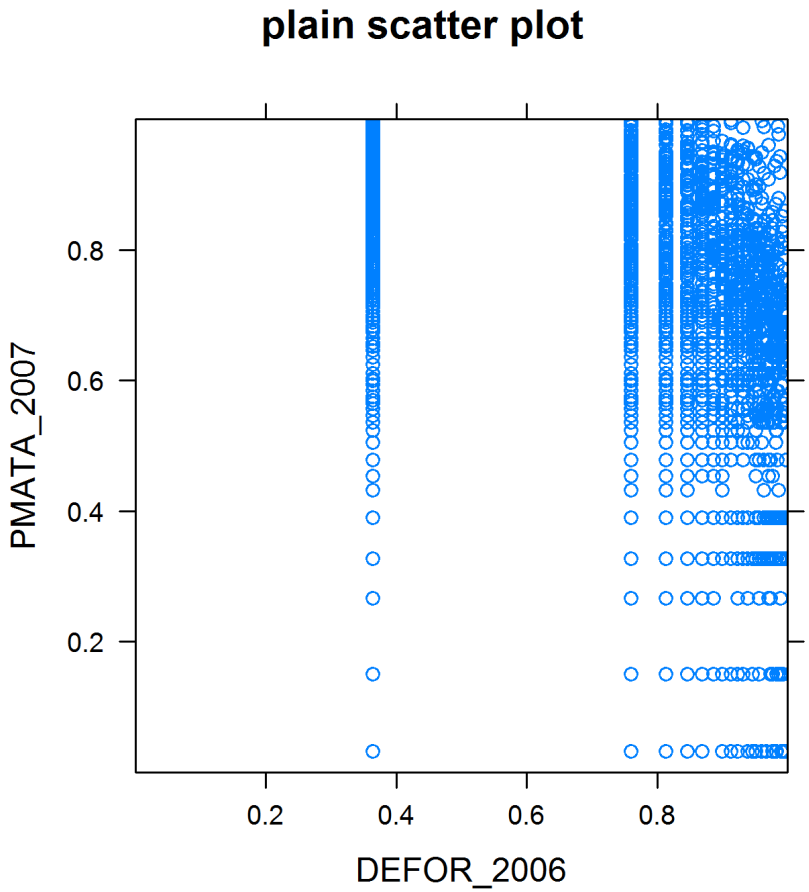
The top right part is rescaled to uniformity and a copula is fitted:



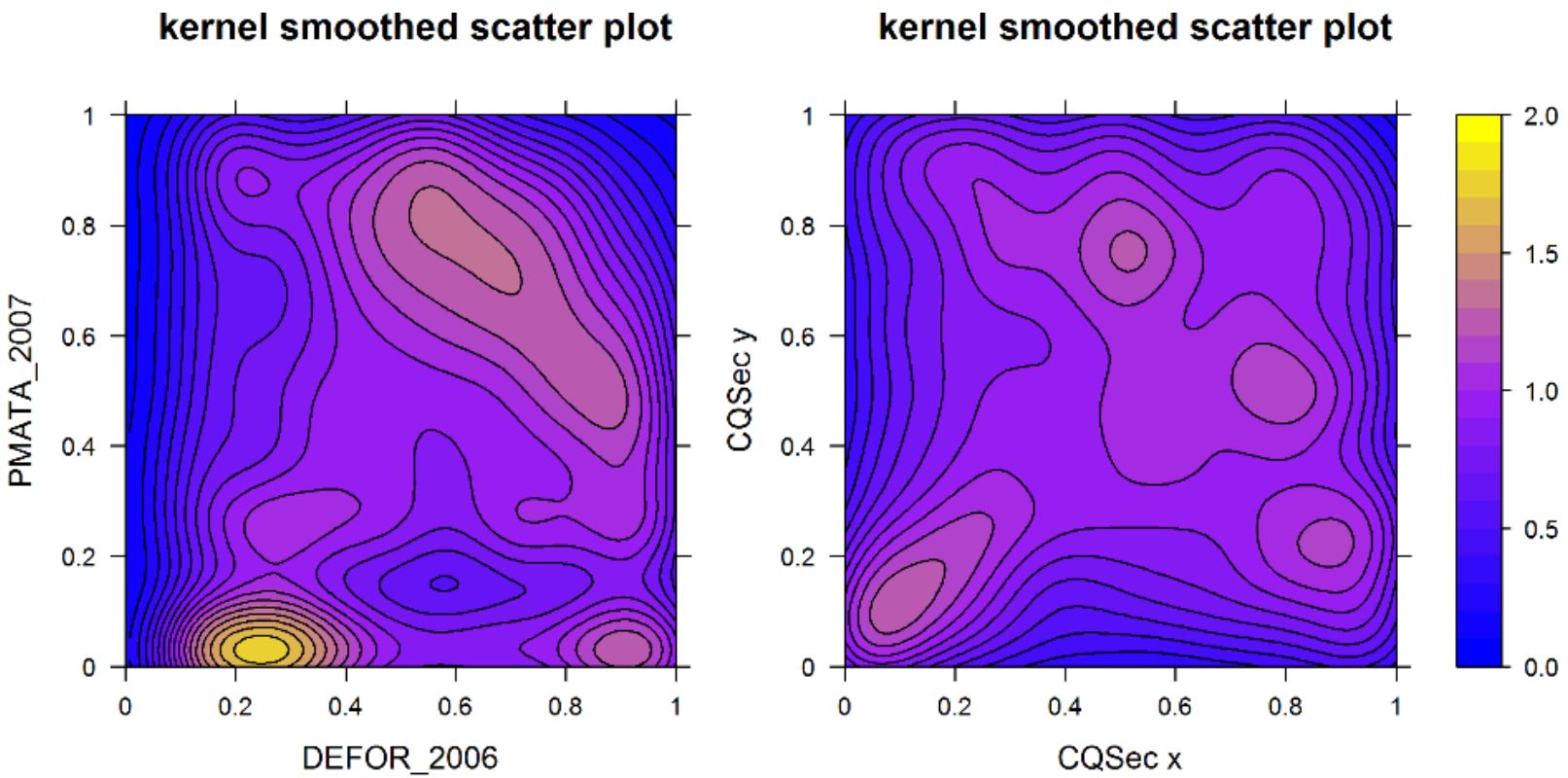
The three other parts are modeled through empirical cumulative distribution functions.

the second pair I

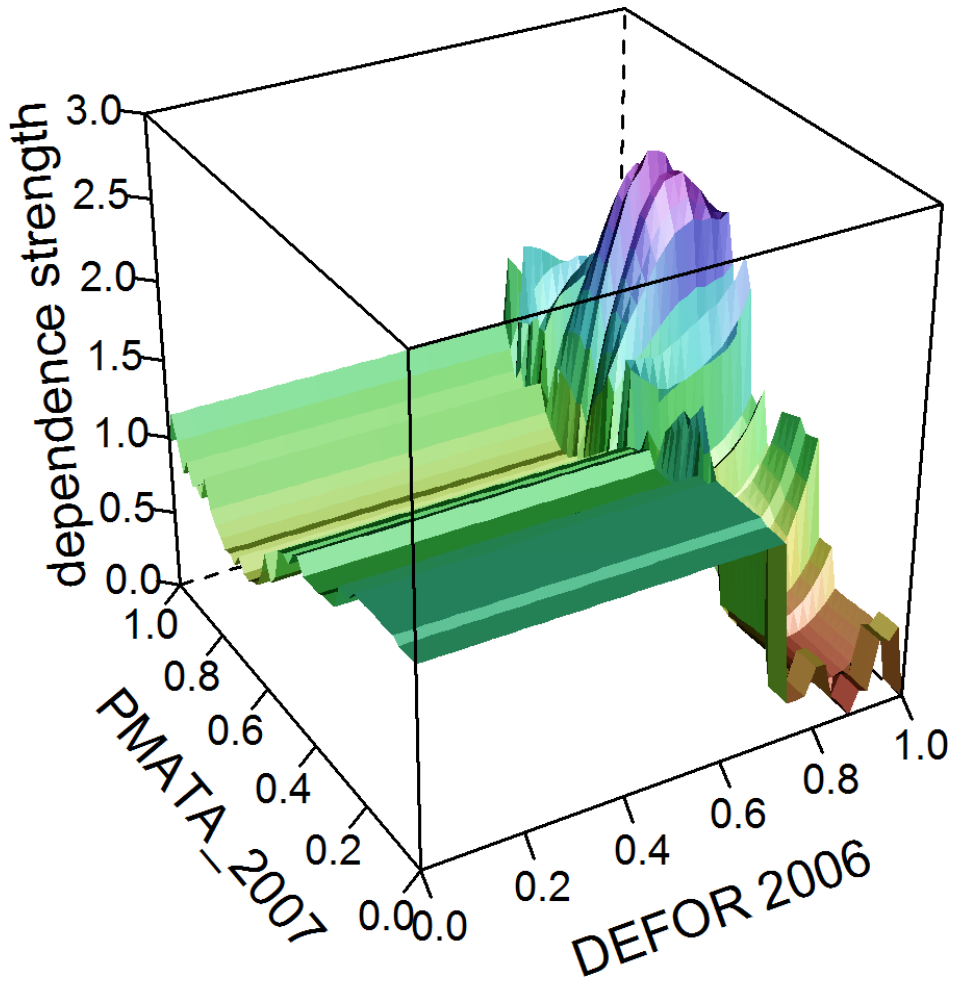
The same procedure is applied to the second pair variables:



the second pair II

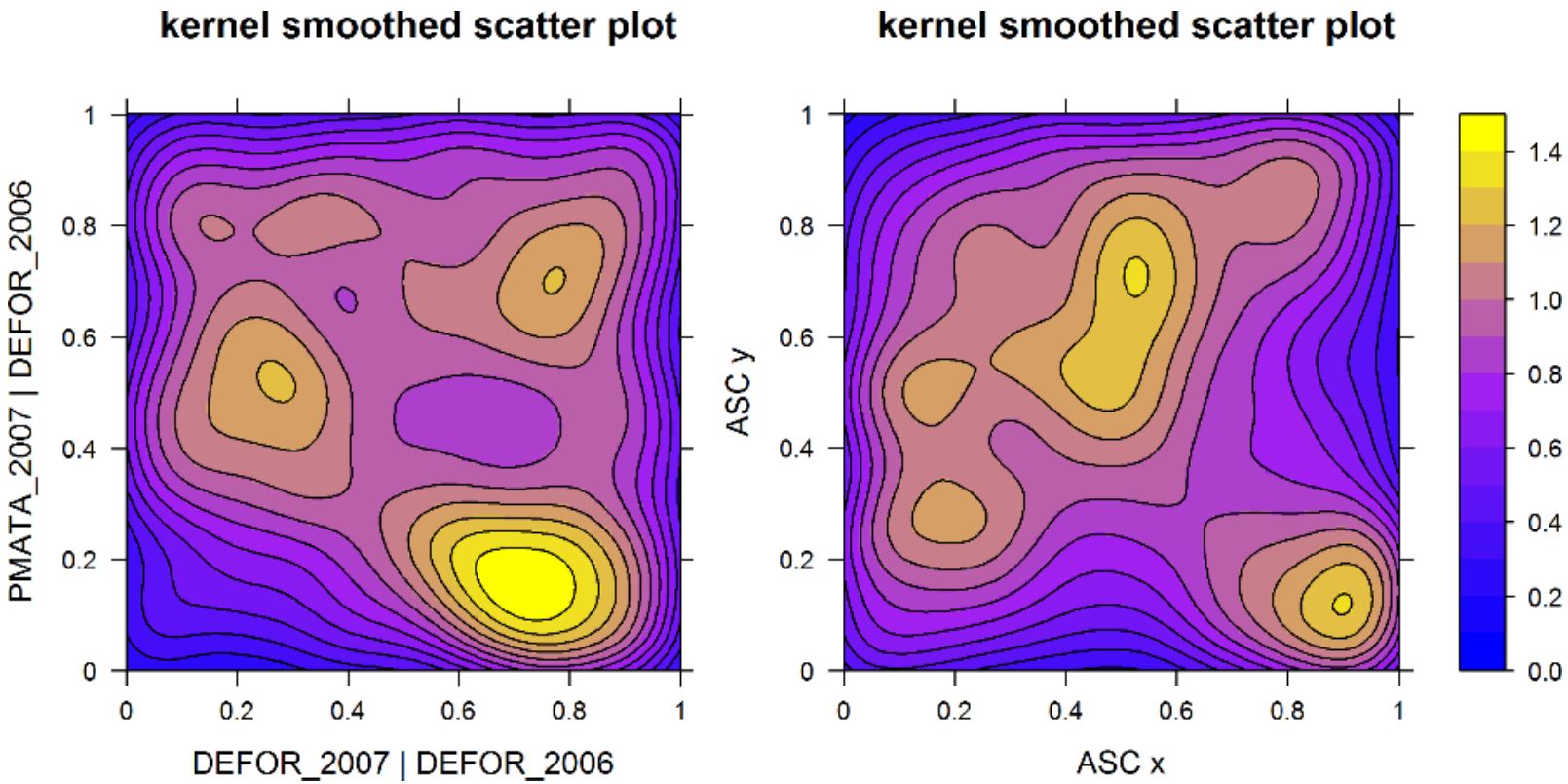


the second pair III

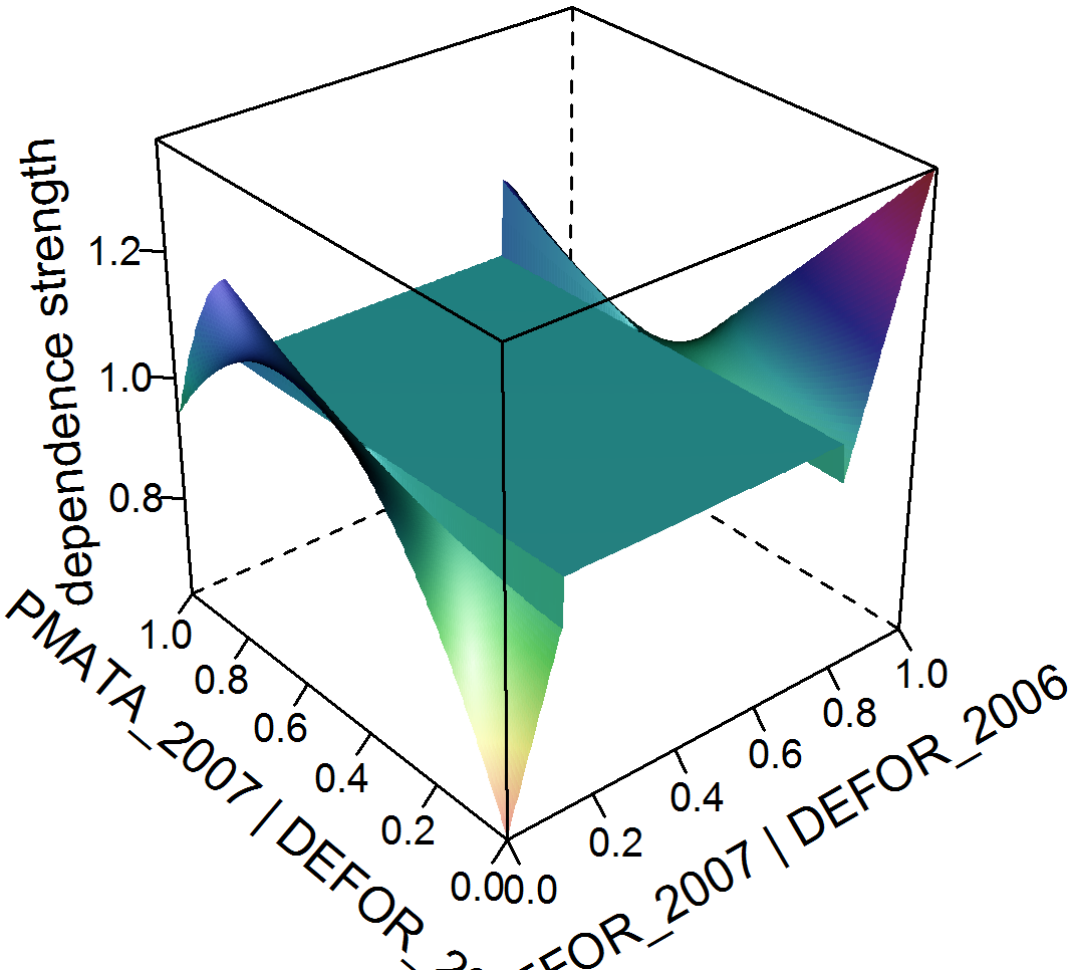


the cutted copula I

After the transformation of the data under the conditional distribution theres is a second value which takes a massive mass.
The remaining copula is:



the cutted copula II



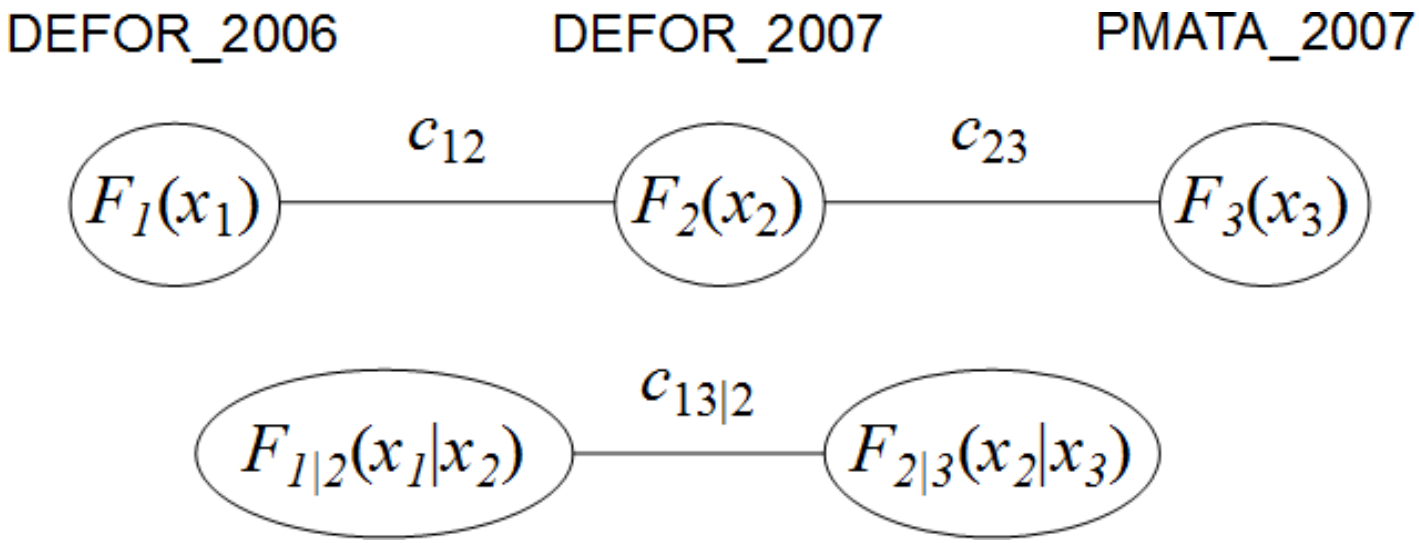
Benedikt Gräler

ifgi
Institute for Geoinformatics
University of Münster

the joined copula

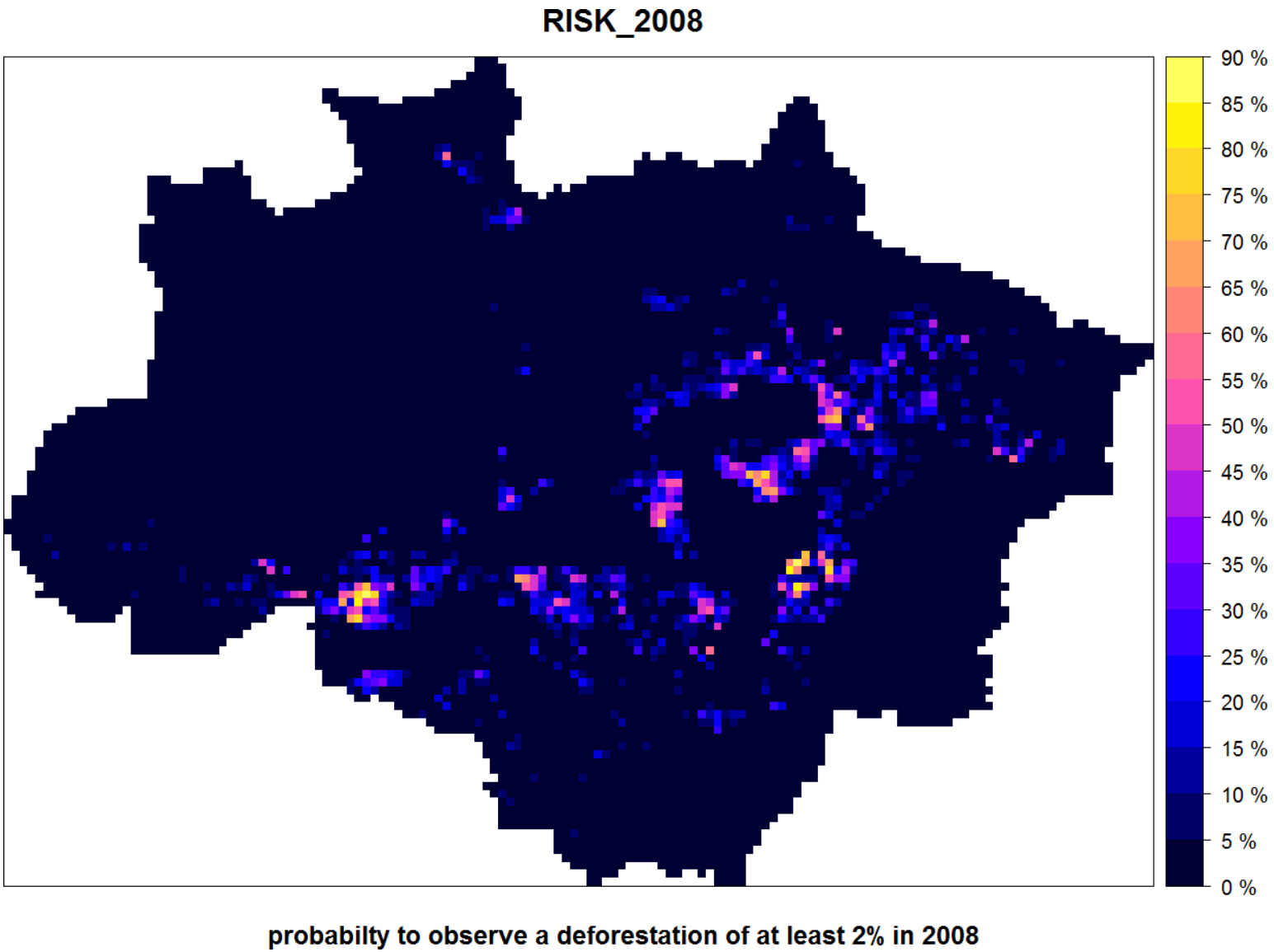
ifgi
Institute for Geoinformatics
University of Münster

After we estimated the three pieces we can put them together:



Risk

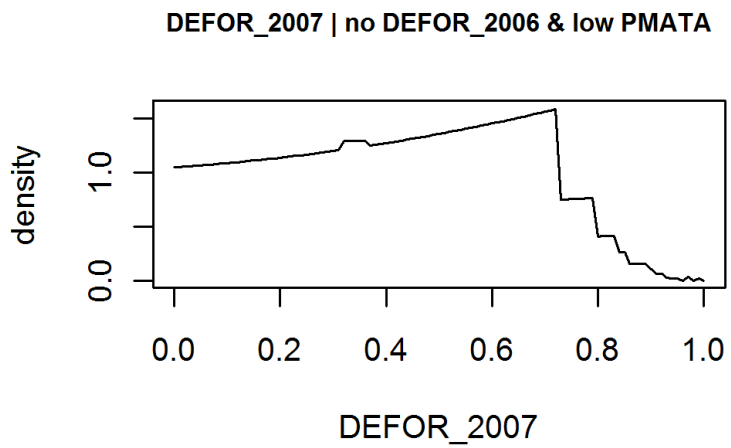
Assuming temporal stationarity lets us calculate a risk map of deforestation for a given threshold

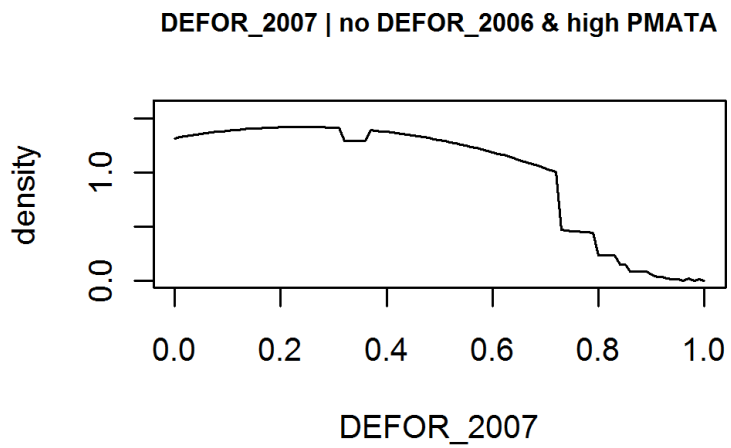


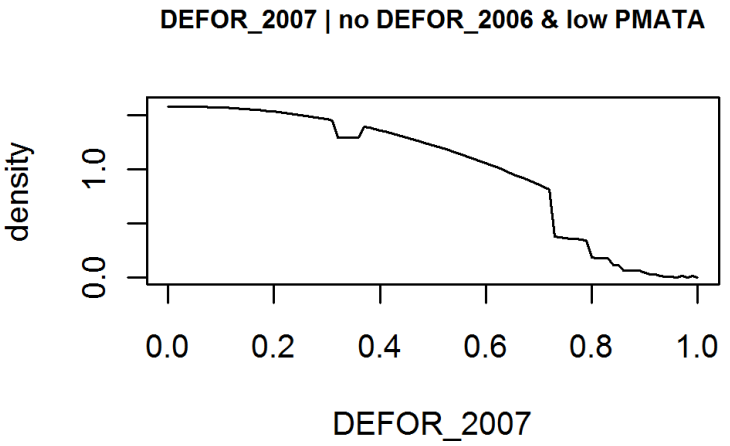
ifgi
Institute for Geoinformatics
University of Münster

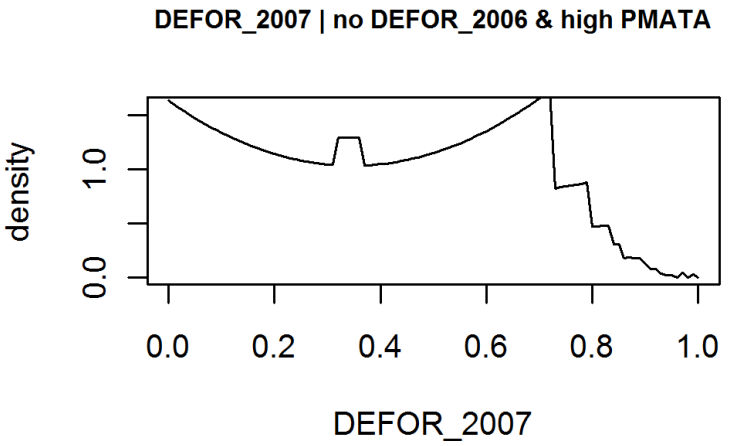
inspecting conditional densities for different copulas

Substituting the CQSec copula (top row) C_{23} with the best Gaussian (bottom row) has a visible impact:









Benedikt Gräler

ifgi
Institute for Geoinformatics
University of Münster

Practical

Choose your own triple of variables of the amazon / meuse / your data and try to estimate a pair copula.

or

Continue with the local neighborhood approach from the last practical to design an interpolation method incorporating the conditional density of the copula.

ifgi
Institute for Geoinformatics
University of Münster

References & further readings I

- Gräler, Benedikt (2009), 'Copulas for Spatio-Temporal Random Fields', Diploma thesis at the Institute of Mathematical Statistics and Institue for Geoinformatics, University of Muenster.
- Gräler, B., H. Kazianka & G. M. de Espindola (2010): "Copulas, a novel approach to model spatial and spatio-temporal dependence". In K. Henneböhl, L. Vinhas, E. Pebesma, & G. Cãmara (Eds.), GIScience for Environmental Change Symposium Proceedings, ifgiprints (Vol. 40, pp. 49-54). Presented at the GIScience for Environmental Change, November 27, 2010, Campos do Jordão (São Paulo), Brazil: AKA Verlag.

ifgi
Institute for Geoinformatics
University of Münster

References & further readings II

- Gräler, Benedikt & Edzer Pebesma (2011): The pair-copula construction for spatial data: a new approach to model spatial dependency. Poster at: Spatial Statistics 2011 - Mapping global change. Enschede, The Netherlands, 23-25 March 2011 (accepted for prersentation and publications in *Procedia Environmental Sciences* by Elsevier).
- Nelsen R. B. (2006): An Introduction to Copulas, 2nd Edition, Springer Science+Buisness, New York