Vine Copulas for Spatial Interpolation

4th Workshop on Vine Copula Distributions and Applications: Vines and Spatial Applications
12.05.2011

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

1

spatial data

Vine Copulas for
Spatial
Interpolation

Benedikt Gräler

ifgi
Institute for Geoinformatics
University of Münster

Spatially distributed data is usually collected at a discrete set of locations in space only.

Maps are often designed to provide information of the full random field $\mathcal{Z}$ at any location $s \in S$ based on some interpolation algorithm.

In geostatistics, we treat the observations as realisations of some spatial random field $\mathcal{Z} : S \to \mathbb{R}$ over some spatial region S.

Tobler's first law of geography

Waldo Tobler (1970)[1]:
"Everything is related to everything else, but near things are more related than distant things."

The most simple idea is the *inverse distance weighted interpolation* that estimates the value at an unobserved location by the linear combination of all measurements weighted by the inverse of their distance to the unobserved location.

---

[1]Tobler, W. (1970): A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46(2): 234-240

Kriging

The most commonly used method (or set of methods) is *kriging*.

Kriged estimates are linear combinations of all values; weights are chosen based on covariances of all pairs of locations (BLUE).

The assumption of stationarity across the region S allows to consider lag classes by distance instead of single locations in space.

Our approach

- using the full distribution of the observed phenomenon of a local neighbourhood

ifgi
Institute for Geoinformatics
University of Münster

Our approach

Vine Copulas for
Spatial
Interpolation

Benedikt Gräler

ifgi
Institute for Geoinformatics
University of Münster

- using the full distribution of the observed phenomenon of a local neighbourhood
- each observed location $s_0 \in S$, is grouped with its four nearest neighbours $\{s_1, s_2, s_3, s_4\} \subset S$ to generate a five dimensional sample: $(\mathbf{X}_0, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4)$

Our approach

Vine Copulas for
Spatial
Interpolation

Benedikt Gräler

ifgi
Institute for Geoinformatics
University of Münster

- using the full distribution of the observed phenomenon of a local neighbourhood
- each observed location $s_0 \in S$, is grouped with its four nearest neighbours $\{s_1, s_2, s_3, s_4\} \subset S$ to generate a five dimensional sample: $(\mathbf{X}_0, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4)$
- an estimate is calculated from the conditional distribution at an unobserved location conditioned under the values of its neighbourhood

Our spatial pair-copula

We decompose the five dimensional distribution into its marginal distribution F (identical for all 5 margins) and a vine copula:
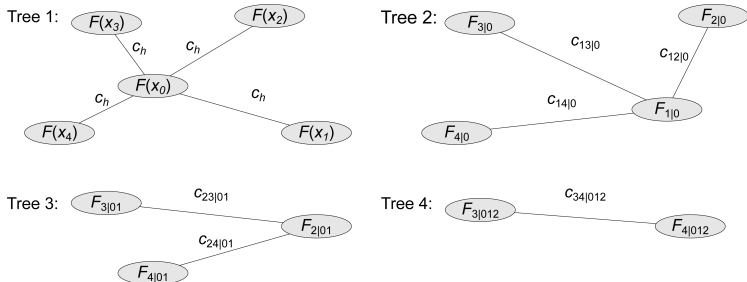

Figure: Modified canonical vine from [1].

Accounting for distance

Vine Copulas for
Spatial
Interpolation

Benedikt Gräler

ifgi
Institute for Geoinformatics
University of Münster

Problem

Solution
Bivariate Spatial
Copula
Spatial Pair-Copula
interpolation
Software

Application
Meuse river
Goodness of fit

Conclusion &
Outlook

References

7

Thinking of pairs of locations we assume . . .

distance has a strong influence on the strength of
dependence

dependence structure is identical for all neighbours, but
might change with distance

stationarity and build k lag classes by distance and
estimate a bivariate copula $c_{j,\tau(h)}(u, v)$ for
each lag class $[0, l_1), [l_1, l_2), \ldots, [l_{k-1}, l_k)$

The function $\tau(h)$ maps distances to a set of estimated
parameters.

Density of the spatial copula

The density of the *bivariate spatial copula* is then given by a convex combination of bivariate copula densities:

$$c_h(u,v) := \begin{cases} c_{1,\tau(h)}(u,v) & ,0 \le h < l_1 \\ (1-\lambda_2)c_{1,\tau(h)}(u,v) + \lambda_2 c_{2,\tau(h)}(u,v) & ,l_1 \le h < l_2 \\ \vdots & \vdots \\ (1-\lambda_k)c_{k-1,\tau(h)}(u,v) + \lambda_k \cdot 1 & ,l_{k-1} \le h < l_k \\ 1 & ,l_k \le h \end{cases}$$

where $\lambda_j := \frac{h - l_{j-1}}{l_j - l_{j-1}}$.

Estimating tree 2 to 4

Vine Copulas for
Spatial
Interpolation

Benedikt Gräler

ifgi
Institute for Geoinformatics
University of Münster

Problem

Solution
Bivariate Spatial
Copula
Spatial Pair-Copula
interpolation
Software

Application
Meuse river
Goodness of fit

Conclusion &
Outlook

References

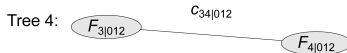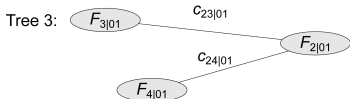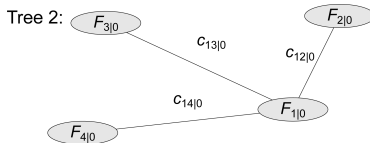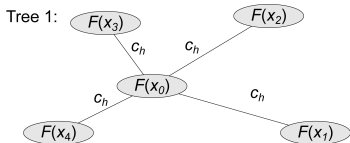The remaining copulas are estimated as usual over the five dimensional sample:

Figure: Modified canonical vine from [1].

Vine Copulas for
Spatial
Interpolation

Benedikt Gräler

ifgi
Institute for Geoinformatics
University of Münster

Problem

Solution
Bivariate Spatial
Copula
Spatial Pair-Copula
interpolation
Software

Application
Meuse river
Goodness of fit

Conclusion &
Outlook

References

The full five dimensional copula density

We get the full density

$$c_h(u_0, u_1, \ldots, u_4)$$
$$= \prod_{i=1}^{4} c_{h_{0i}}(u_0, u_i) \cdot \prod_{j=2}^{4} c_{1j|0}\big(F_{1|0}, F_{j|0}\big)$$
$$\cdot \prod_{m=3}^{4} c_{2m|01}\big(F_{2|01}, F_{m|01}\big) \cdot c_{34|012}\big(F_{3|012}, F_{4|012}\big)$$

and the conditional density

$$c_h(u_0|u_1, \ldots, u_4) := \frac{c_h(u_0, u_1, \ldots, u_4)}{\int_0^1 c_h(v, u_1, \ldots, u_4)dv}$$

where $h := (h_{01}, h_{02}, h_{03}, h_{04})$ is the vector of separating distances and $u_i := F(x_i)$.

Spatial Pair-Copula interpolation

Vine Copulas for Spatial Interpolation

Benedikt Gräler

ifgi
Institute for Geoinformatics
University of Münster

Problem

Solution
Bivariate Spatial Copula
Spatial Pair-Copula interpolation
Software

Application
Meuse river
Goodness of fit

Conclusion & Outlook

References

The estimate can be obtained as the expected value

$$\hat{x}_0 = \int_0^1 F^{-1}(v) \cdot c_h\big(v|u_1, \dots u_4\big) dv$$

or by calculating the median

$$\hat{x}_0 = F^{-1}\big(C_h^{-1}(0.5|u_1, \dots u_4)\big)$$

where $h := (h_{01}, h_{02}, h_{03}, h_{04})$ and $u_i := F(x_i)$ as before.

R-package spcopula

ifgi
Institute for Geoinformatics
University of Münster

The developed methods are implemented as R-scripts and will be bundled in the package *spcopula*.

It extends and combines the R-packages sp and copula.

Look for the project spcopula at R-Forge.

The Meuse river I

data measurements of four heavy metals in soil samples at a small section of the river Meuse at the Dutch-Belgian border

task interpolation of the zinc measurements for the whole study area

margins Fréchet distribution

estimation Kendall's tau estimator, comparison of likelihoods

bivariate spatial copula mixture of Gumbel and Clayton copulas

spatial pair-copula mainly from the Gumbel family; one copula with cubic-quadratic-sections [2]

The Meuse river II

The bivariate spatial copula for the Meuse data set:
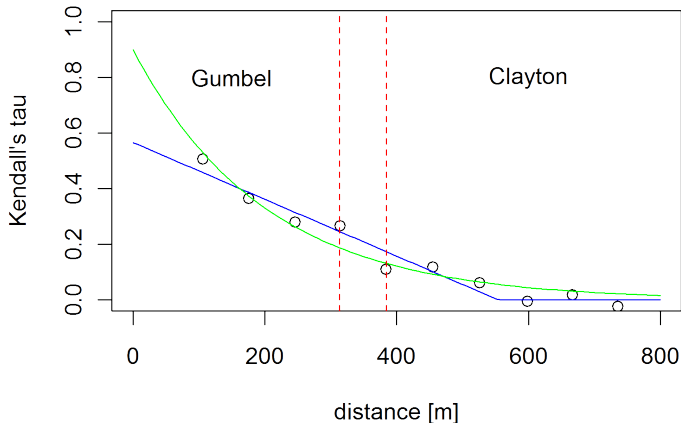
The Meuse river III

ifgi
Institute for Geoinformatics
University of Münster

zinc predictions

90% conf. band width (zinc)

comparing likelihoods

Vine Copulas for
Spatial
Interpolation

Benedikt Gräler

ifgi
Institute for Geoinformatics
University of Münster

Problem

Solution
Bivariate Spatial
Copula
Spatial Pair-Copula
interpolation
Software

Application
Meuse river
Goodness of fit
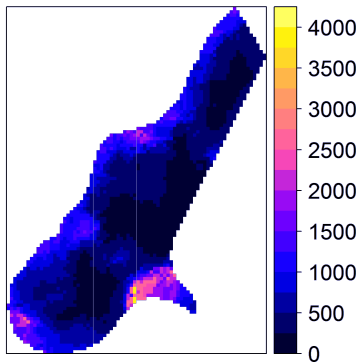
Conclusion &
Outlook

References

5-dim copula	log-likelihood
spatial pair-copula	203
. . . with rev. distances	187
Gaussian	174
Gumbel	152
Clayton	138
Frank	135

Cross validation

estimates vs. measurements

The most popular spatial interpolation

Vine Copulas for
Spatial
Interpolation

Benedikt Gräler

ifgi
Institute for Geoinformatics
University of Münster

pair-copula interpolation vs. kriging[2]:

      bias could be reduced by a factor of 2

---

[2]using the R-package gstat

The most popular spatial interpolation

Vine Copulas for
Spatial
Interpolation

Benedikt Gräler

ifgi
Institute for Geoinformatics
University of Münster

pair-copula interpolation vs. kriging[2]:

  bias could be reduced by a factor of 2

  RMSE was slightly higher, but appeared to be smaller
  for some random sub samples from the study
  area

---

[2]using the R-package gstat

The most popular spatial interpolation

Vine Copulas for Spatial Interpolation

Benedikt Gräler

ifgi
Institute for Geoinformatics
University of Münster

Problem

Solution
Bivariate Spatial Copula
Spatial Pair-Copula interpolation
Software

Application
Meuse river
Goodness of fit

Conclusion & Outlook

References

pair-copula interpolation vs. kriging[2]:

bias could be reduced by a factor of 2

RMSE was slightly higher, but appeared to be smaller for some random sub samples from the study area

probabilistic advantage sophisticated uncertainty analysis, drawing random samples, . . .

---

[2]using the R-package gstat

Further extensions

Vine Copulas for
Spatial
Interpolation

Benedikt Gräler

ifgi
Institute for Geoinformatics
University of Münster

- including covariates

Further extensions

Vine Copulas for
Spatial
Interpolation

Benedikt Gräler

ifgi
Institute for Geoinformatics
University of Münster

- including covariates
- complex neighbourhoods
  (e.g. closest neighbours in the north, east, south and west)

Further extensions

Vine Copulas for
Spatial
Interpolation

Benedikt Gräler

ifgi
Institute for Geoinformatics
University of Münster

- including covariates
- complex neighbourhoods
  (e.g. closest neighbours in the north, east, south and west)
- accounting as well for direction

Further extensions

- including covariates
- complex neighbourhoods
  (e.g. closest neighbours in the north, east, south and west)
- accounting as well for direction
- temporal dependence
  (e.g. 2-dim convex combination of bivariate copulas)

Further extensions

- including covariates
- complex neighbourhoods
  (e.g. closest neighbours in the north, east, south and
  west)
- accounting as well for direction
- temporal dependence
  (e.g. 2-dim convex combination of bivariate copulas)
- include further copula families

Further extensions

- including covariates
- complex neighbourhoods
  (e.g. closest neighbours in the north, east, south and
  west)
- accounting as well for direction
- temporal dependence
  (e.g. 2-dim convex combination of bivariate copulas)
- include further copula families
- improve performance

Vine Copulas for Spatial Interpolation

Benedikt Gräler

ifgi
Institute for Geoinformatics
University of Münster

Problem

Solution
Bivariate Spatial Copula
Spatial Pair-Copula interpolation
Software

Application
Meuse river
Goodness of fit

Conclusion & Outlook

References

📕 Aas, Kjersti, Claudia Czado, Arnoldo Frigessi & Henrik Bakken (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics, 44*, 182 - 198.

📕 Nelsen, R. (2006). *An Introduction to Copulas*. 2nd ed. New York: Springer.