Chapter 1

Copulas in Spatial Statistics

Summer School GEOSTAT 2014,
Spatio-Temporal Geostatistics,
2014-06-19

Benedikt Gräler

http://ifgi.de/graeler
Institute for Geoinformatics
University of Muenster
What if the world happens to be non-Gaussian?

<table>
<thead>
<tr>
<th></th>
<th>Gumbel – 185</th>
<th>Gaussian – 185</th>
<th>Clayton – 185</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>150</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gumbel – 125</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>140</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>130</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>120</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gumbel – 95</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>110</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>105</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Copulas

Bivariate Spatial Copula
Spatial Vine Copula
Software

Application to Nuclear Radiation
Fitment
Goodness of Fit

Conclusion & Outlook

References
Bivariate Copulas I

Copulas allow to model dependencies much more detailed than a typical correlation value.

Instead of a single value, a full distribution is fitted describing dependence.
Bivariate Copulas II
See the copulatheque for further interactive examples.
Behind the scenes - Sklar’s Theorem

Every bivariate distribution $H$ is composed out of some copula $C$ and marginal distributions $F_1$ and $F_2$:

$$H(x, y) = C(F_1(x), F_2(y))$$
Why is this useful?

Sklar’s theorem allows us to model any multivariate distribution in two steps:

1. find marginal distribution functions using your favourite estimation technique that suite the data
2. find a copula that describes the dependence

This allows for a huge flexibility and a clear outline how to proceed.

See the interactive copulatheque on my website for more copula families.
Accounting for distance

Thinking of pairs of locations \((s_i, s_j)\) 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 \(k\) bins by spatial distance and estimate a bivariate copula \(c_j(u, v)\) for each bin \([0, l_1), [l_1, l_2), \ldots, [l_{k-1}, l_k)\)
Accounting for distance

Thinking of pairs of locations \((s_i, s_j)\) 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 \(k\) bins by spatial distance and estimate a bivariate copula \(c_j(u, v)\) for each bin \([0, l_1), [l_1, l_2), \ldots, [l_{k-1}, l_k)\)
Accounting for distance

Thinking of pairs of locations \((s_i, s_j)\) 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 \(k\) bins by spatial distance and estimate a bivariate copula \(c_j(u, v)\) for each bin \([0, l_1), [l_1, l_2), \ldots, [l_{k-1}, l_k)\)
Density of the bivariate spatial copula

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

\[
c_h(u, v) := \begin{cases} 
  c_1(u, v) & \text{, } 0 \leq h < l_1 \\
  (1 - \lambda_2)c_1(u, v) + \lambda_2 c_2(u, v) & \text{, } l_1 \leq h < l_2 \\
  \vdots & \vdots \\
  (1 - \lambda_k)c_{k-1}(u, v) + \lambda_k \cdot 1 & \text{, } l_{k-1} \leq h < l_k \\
  1 & \text{, } l_k \leq h
\end{cases}
\]

where \( \lambda_j := \frac{h - l_{j-1}}{l_j - l_{j-1}} \). Each tree has its own bivariate spatial copula where distance \( h \) relates the involved pairs of locations.
The spatial neighbourhood

Spatial tree 1:
- $Z(s_3)$
- $Z(s_2)$
- $Z(s_0)$
- $Z(s_4)$

Spatial tree 2:
- $F_{3|0}$
- $F_{2|0}$

Spatial tree 3:
- $F_{3|1}$
- $F_{2|0}$

Spatial tree 4:
- $F_{3|02}$
- $F_{4|01}$
- $F_{4|02}$

Copulas
Bivariate Spatial Copula
Spatial Vine Copula
Software
Application to Nuclear Radiation
Fitment
Goodness of Fit
Conclusion & Outlook
References
The upper trees

Using the bivariate spatial copula on the first tree, the sample conditioned on \( s_0 \) is obtained.

The next bivariate spatial copula uses the distances between locations \( (s_1, s_2), (s_1, s_3), \ldots, (s_1, s_d) \).

A spatial binning allows to estimate the next bivariate spatial copula to generate the sample conditioned on \( s_0 \) and \( s_1 \).
The full density

We get the full copula density as a product of all involved bivariate densities:

\[ c_h(u_0, \ldots, u_d) \]
\[ = \prod_{i=1}^{d} c_{h_0(i)}(u_0, u_i) \cdot \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{h_j(j+i)}(u_j|0, \ldots, j-1, u_j+i|0, \ldots, j-1) \]

where \( u_i = F_i(Z(s_i)) \) for \( 0 \leq i \leq d \) and

\[ u_{j+i|0,\ldots,j-1} = F_{h_{j-1}(j+i)}(u_{j+i}|u_0, \ldots, u_j-1) \]
\[ = \frac{\partial C_{h_{j-1}(j+i)}(u_{j-1}|0, \ldots, j-2, u_{j+i}|0, \ldots, j-2)}{\partial u_{j-1}|0, \ldots, j-2} \]
Spatial vine copula interpolation

The estimate can be obtained as the expected value

$$\hat{Z}_m(s_0) = \int_{[0,1]} F^{-1}(u) \ c_h(u|u_1, \ldots, u_d) \ du$$

or by calculating any percentile $p$ (i.e. the median)

$$\hat{Z}_p(s_0) = F^{-1}(C_h^{-1}(p|u_1, \ldots, u_d))$$

with the conditional density

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

and $u_i = F(Z(s_i))$ as before.
R-package spcopula

The developed methods are implemented as R-scripts and are bundled in the package spcopula available at R-Forge. The package spcopula extends and combines the R-packages VineCopula, spacetime and copula.
Simulated nuclear radiation

The new method is applied to simulated nuclear radiation data mimicking an emergency scenario.

The data has been generated for the spatial interpolation comparison 2004 (SIC2004, [? ]) with 200 data and 808 validation locations.

To better approximate stationarity, a quadratic trend surface has been fitted excluding the (two) extremes.

This has been published in Journal of Spatial Statistics [1].
The trend surface

<table>
<thead>
<tr>
<th>trendSurf</th>
<th>IDW</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- 60
- 80
- 100
- 120
- 140
- 160
- 180
- 200
The marginal distribution

**Figure:** Histogram of the "observed" radiation values.

The empirical marginal distribution function has been used.
The bivariate spatial copulas I

Strength of dependence on copula scale

![Diagram showing the strength of dependence on copula scale for different spatial scales.](image)
The bivariate spatial copulas II

- Kendall's tau
- distance [m]
- spatial copula \( h_0 \)
- spatial copula \( h_1 \)
- spatial copula \( h_2 \)
- spatial copula \( h_3 \)
Interpolated grid
Validation data set

808 locations have been hold back to assess the prediction quality.

<table>
<thead>
<tr>
<th>approach</th>
<th>MAE</th>
<th>RMSE</th>
<th>ME</th>
<th>COR</th>
</tr>
</thead>
<tbody>
<tr>
<td>spatial vine copula</td>
<td>14.5</td>
<td>67.6</td>
<td>-6.1</td>
<td>0.60</td>
</tr>
<tr>
<td>TG log-kriging</td>
<td>20.8</td>
<td>78.2</td>
<td>-2.1</td>
<td>0.39</td>
</tr>
<tr>
<td>residual kriging</td>
<td>21.1</td>
<td>75.6</td>
<td>5.2</td>
<td>0.43</td>
</tr>
<tr>
<td>best one in SIC2004</td>
<td>14.9</td>
<td>45.5</td>
<td>-0.5</td>
<td>0.84</td>
</tr>
</tbody>
</table>
Reproduction of margins

radiation [nSv/h]

- emergency, scen.
- pure, emp.
- 4 sp. trees, emp.
- 1 sp. tree, emp.
- sp. Gaussian, emp.
- pure, PoT
- 4 sp. trees, PoT
- 1 sp. tree, PoT
- sp. Gaussian, PoT
- TG log-kriging
- residual kriging
Uncertainty assessment

Copulas in Spatial Statistics
Benedikt Gräler

1.23

Copulas
Bivariate Spatial Copula
Spatial Vine Copula
Software
Application to Nuclear Radiation
Fitment
Goodness of Fit
Conclusion & Outlook
References
Simulation

Predicting random quantiles from the spatial vine copula.
Benefits

richer flexibility due to the various dependence structures
asymmetric dependence structures become possible (temporal direction)
probabilistic advantage sophisticated uncertainty analysis, drawing random samples, ...
Benefits

richer flexibility due to the various dependence structures
asymmetric dependence structures become possible
(temporal direction)

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

richer flexibility due to the various dependence structures
asymmetric dependence structures become possible
(temporal direction)
probabilistic advantage sophisticated uncertainty analysis,
drawing random samples, ...
Further extensions

- including covariates (e.g. altitude, population, EMEP, . . .)
- flexible/complex neighbourhoods (e.g. by spatial direction, . . .)
- larger neighbourhoods possibly using vine truncation techniques
- include further copula families
- improve performance
Further extensions

- including covariates (e.g. altitude, population, EMEP, ...)
- flexible/complex neighbourhoods (e.g. by spatial direction, ...)
- larger neighbourhoods possibly using vine truncation techniques
- include further copula families
- improve performance
Further extensions

- including covariates (e.g. altitude, population, EMEP, ...)
- flexible/complex neighbourhoods (e.g. by spatial direction, ...)
- larger neighbourhoods possibly using vine truncation techniques
  - include further copula families
  - improve performance
Further extensions

- including covariates (e.g. altitude, population, EMEP, ...)
- flexible/complex neighbourhoods (e.g. by spatial direction, ...)
- larger neighbourhoods possibly using vine truncation techniques
- include further copula families
- improve performance
Further extensions

- including covariates
  (e.g. altitude, population, EMEP, . . . )
- flexible/complex neighbourhoods
  (e.g. by spatial direction, . . . )
- larger neighbourhoods possibly using vine truncation techniques
- include further copula families
- improve performance
References I