

The spcopula R-package: Modelling Spatial and Spatio-Temporal Dependence with Copulas

Spatial Statistics 2013, S2.2

6th June 2013

Benedikt Gräler
Institute for Geoinformatics
University of Münster

<http://ifgi.uni-muenster.de/graeler>

The spcopula
R-package:
Modelling Spatial
and
Spatio-Temporal
Dependence with
Copulas

Benedikt Gräler



Problem

Solution

Vine Copulas
Bivariate
Spatio-Temporal
Copulas
Spatio-Temporal Vine
Copulas
Software

Application

Daily Mean PM₁₀

Conclusion &
Outlook

References

Typically, spatial/spatio-temporal data is given at a set of discrete locations $s_i \in S$ (and time steps $t_j \in T$).

We desire a full spatial/spatio-temporal random field Z modelling the process at any location s in space/ (s, t) in space and time.

We will look at daily mean fine dust concentrations across Europe (PM_{10}).

In the following, we will refer to Z as a spatio-temporal random field (including the spatial case).

Problem

Solution

Vine Copulas
Bivariate
Spatio-Temporal
Copulas
Spatio-Temporal Vine
Copulas
Software

Application

Daily Mean PM_{10}

Conclusion & Outlook

References

Copulas describe the dependence structure between the margins of a multivariate distribution. Sklar's Theorem states:

$$H(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d))$$

where H is any multivariate CDF, $F_1, \dots, F_d$ are the corresponding marginal univariate CDFs and C is a suitable *copula* (uniquely determined in a continuous setting).

Since $F_i(X_i) \sim U(0, 1)$, copulas can be thought of as CDFs on the unit (hyper-)cube.

Problem

Solution

Vine Copulas

Bivariate
Spatio-Temporal
Copulas
Spatio-Temporal Vine
Copulas
Software

Application

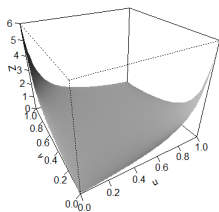
Daily Mean PM₁₀

Conclusion &
Outlook

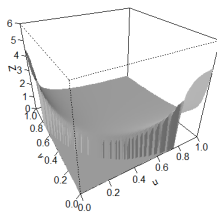
References

Beyond Gaussian dependence structure

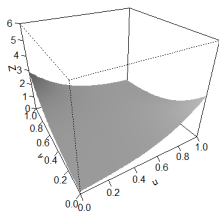
Gaussian



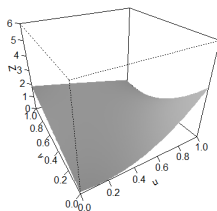
Clayton



Frank



asymmetric



The spcopia
R-package:
Modelling Spatial
and
Spatio-Temporal
Dependence with
Copulas

Benedikt Gräler



Problem

Solution

Vine Copulas

Bivariate
Spatio-Temporal
Copulas
Spatio-Temporal Vine
Copulas
Software

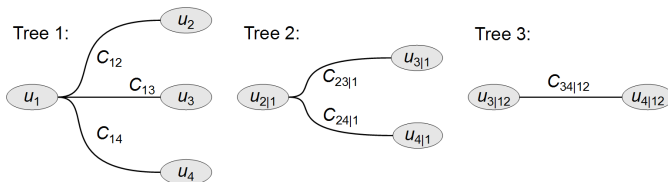
Application

Daily Mean PM₁₀

Conclusion &
Outlook

References

- Bivariate copulas are pretty well understood and rather easy to estimate.
- Unfortunately, most bivariate families do not nicely extend to a multivariate setting or lack the necessary flexibility.
- *Vine Copulas* allow to approximate multivariate copulas by mixing (conditional) bivariate copulas following a vine decomposition [ACFB09, BC02].



Problem

Solution

Vine Copulas

Bivariate
Spatio-Temporal
Copulas
Spatio-Temporal Vine
Copulas
Software

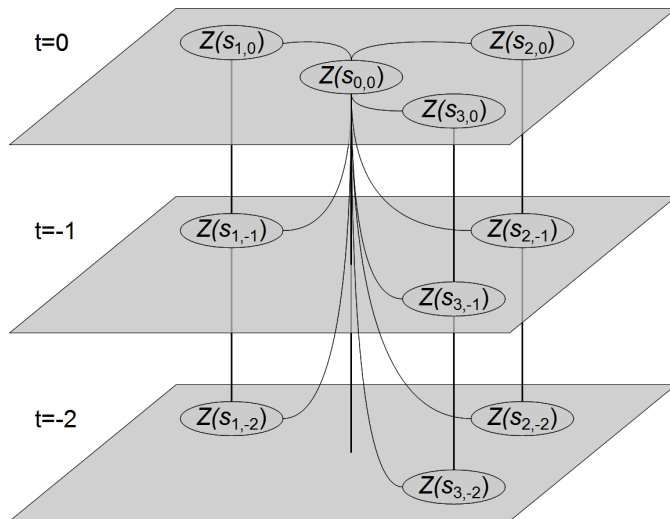
Application

Daily Mean PM₁₀

Conclusion & Outlook

References

The spatio-temporal neighbourhood - the first tree



Problem

Solution

Vine Copulas

- Bivariate
- Spatio-Temporal
- Copulas
- Spatio-Temporal Vine
- Copulas
- Software

Application

Daily Mean PM₁₀

Conclusion & Outlook

References

The distribution of local neighbourhoods is decomposed into marginal distributions F_i and a *spatio-temporal vine copula*:

- The first tree is modelled as bivariate spatio-temporal copulas accounting for spatial and temporal distances.
- Remaining trees are modelled from a wide set of "classical" bivariate copulas as a vine or truncated vine.

Problem

Solution

Vine Copulas

Bivariate
Spatio-Temporal
Copulas
Spatio-Temporal Vine
Copulas
Software

Application

Daily Mean PM₁₀

Conclusion &
Outlook

References

Thinking of pairs of spatio-temporal locations
 $((s_1, t_1), (s_2, t_2))$ we assume ...

distance has a strong influence on the strength of
dependence

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

stationarity and build k bins by distance to estimate a
bivariate copula $c_{j,h}(u, v)$ for all spatial bins
 $[0, l_1), [l_1, l_2), \dots, [l_{k-1}, l_k)$ per temporal lag.

Problem

Solution

Vine Copulas

Bivariate
Spatio-Temporal
Copulas

Spatio-Temporal Vine
Copulas
Software

Application

Daily Mean PM₁₀

Conclusion &
Outlook

References

Density of the spatial and spatio-temporal copula

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

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

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

The density of the *bivariate spatio-temporal copula* $c_{h,\Delta}(u, v)$ is then given by a convex combination of bivariate spatial copula densities in an analogous manner.

Problem

Solution

Vine Copulas

Bivariate
Spatio-Temporal
Copulas

Spatio-Temporal Vine
Copulas
Software

Application

Daily Mean PM₁₀

Conclusion & Outlook

References

The full density

The remaining copulas $c_{j,j+i|0,\dots,j-1}$ are estimated over the conditional sample.

We get the full (here 10-dim) spatio-temporal vine copula density as a product of all involved bivariate densities:

$$\begin{aligned} & c_{\mathbf{h},\Delta}(u_0, \dots, u_9) \\ &= \prod_{i=1}^9 c_{h,\Delta}(u_0, u_i) \cdot \prod_{j=1}^{9-1} \prod_{i=1}^{9-j} c_{j,j+i|0,\dots,j-1}(u_{j|0,\dots,j-1}, u_{j+i|0,\dots,j-1}) \end{aligned}$$

Problem

Solution

Vine Copulas

Bivariate
Spatio-Temporal
Copulas

Spatio-Temporal Vine
Copulas

Software

Application

Daily Mean PM₁₀

Conclusion & Outlook

References

The estimate can be obtained as the expected value

$$\hat{Z}_m(s_0, t_0) = \int_{[0,1]} F^{-1}(u) c_{\mathbf{h}, \Delta}(u|u_1, \dots, u_d) du$$

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

$$\hat{Z}_p(s_0, t_0) = F^{-1}(C_{\mathbf{h}, \Delta}^{-1}(p|u_1, \dots, u_d))$$

with the conditional density

$$c_{\mathbf{h}, \Delta}(u|u_1, \dots, u_d) := \frac{c_{\mathbf{h}, \Delta}(u, u_1, \dots, u_d)}{\int_0^1 c_{\mathbf{h}, \Delta}(v, u_1, \dots, u_d) dv}$$

and $u_i = F_i(Z(s_i, t_i))$.

Problem

Solution

Vine Copulas

Bivariate
Spatio-Temporal
Copulas

Spatio-Temporal Vine
Copulas

Software

Application

Daily Mean PM₁₀

Conclusion &
Outlook

References

The developed methods are implemented in R and are available as package [spcopula](#) at R-Forge.

The package spcopula extends and combines the R-packages [VineCopula](#), [spacetime](#) and [copula](#).

Problem

Solution

- Vine Copulas
- Bivariate
- Spatio-Temporal
- Copulas
- Spatio-Temporal Vine
- Copulas

Software

Application

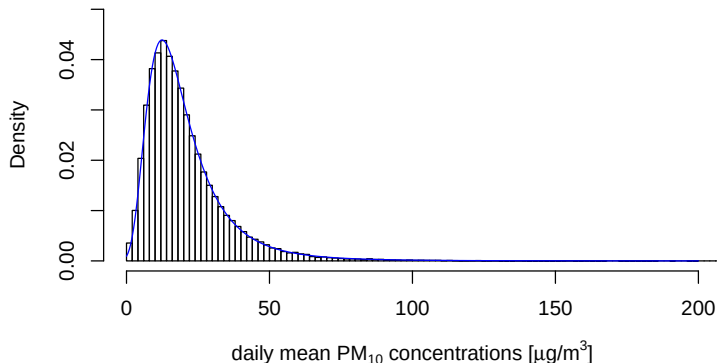
Daily Mean PM₁₀

Conclusion & Outlook

References

Daily mean PM_{10} concentrations across Europe

Daily mean PM_{10} concentrations observed at 194 rural background stations across Europe for the year 2005.



Problem

Solution

- Vine Copulas
- Bivariate
- Spatio-Temporal
- Copulas
- Spatio-Temporal Vine
- Copulas
- Software

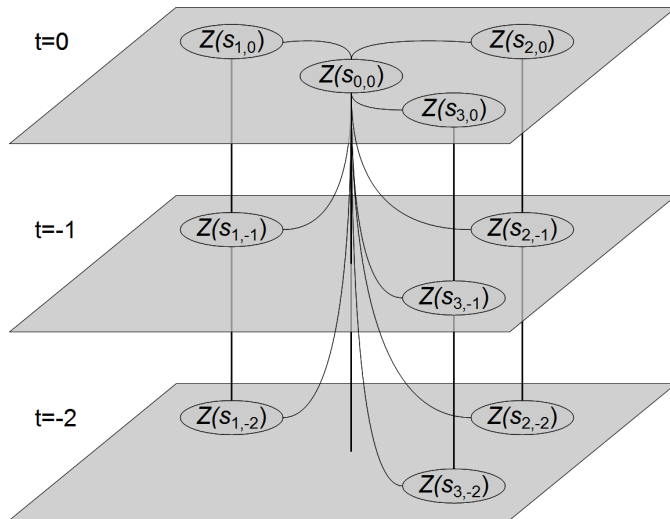
Application

Daily Mean PM_{10}

Conclusion & Outlook

References

The spatio-temporal neighbourhood



Problem

Solution

Vine Copulas
Bivariate
Spatio-Temporal
Copulas
Spatio-Temporal Vine
Copulas
Software

Application

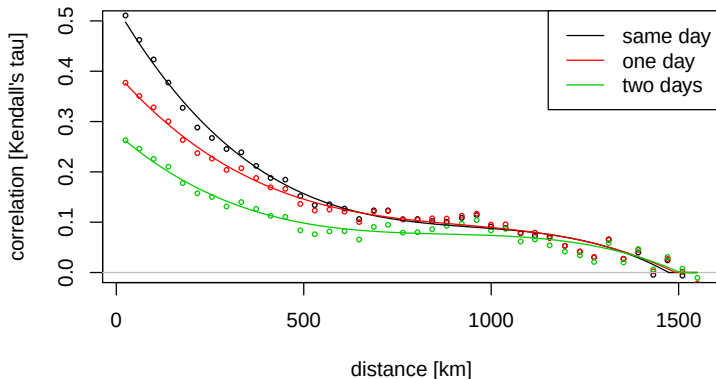
Daily Mean PM₁₀

Conclusion & Outlook

References

The bivariate spatio-temporal copula I

```
> stBins <- calcStBins(EU_RB_2005,"rtPM10", nbins=40,  
+                       t.lags=-(0:2), instances=NA,  
+                       cor.method="fasttau", plot=F)  
> calcKtau <- fitCorFun(stBins,c(3,3,3))
```



Problem

Solution

Vine Copulas
Bivariate
Spatio-Temporal
Copulas
Spatio-Temporal Vine
Copulas
Software

Application

Daily Mean PM₁₀

Conclusion & Outlook

References

The bivariate spatio-temporal copula II

```
> loglikTau <- list()
> for(j in 1:3) {
+   tmpBins <- ... # j-th subset of stBins
+   loglikTau[[j]] <- loglikByCopulasLags(tmpBins, families,
+                                         calcKTau[[j]])
+ }
```

The following families achieve the highest log-likelihood:

distance	25	61	99	139	177	216	255	294	334	373	412	451	491	529	569
$\Delta = 0$	t	F	t	F	F	F	F	F	F	F	F	F	F	F	F
$\Delta = -1$	F	F	F	F	F	F	F	F	F	F	F	F	F	F	A
$\Delta = -2$	F	F	F	F	F	F	A	A	A	A	A	A	A	A	A

Problem

Solution

Vine Copulas
Bivariate
Spatio-Temporal
Copulas
Spatio-Temporal Vine
Copulas
Software

Application

Daily Mean PM₁₀

Conclusion & Outlook

References

The bivariate spatio-temporal copula III

Pick the best fitting copulas:

```
> stConvCOP <- stCOPula(components = listCops,  
+                         distances = listDists,  
+                         t.lags=c(0,-1,-2))
```

The spcopula
R-package:
Modelling Spatial
and
Spatio-Temporal
Dependence with
Copulas

Benedikt Gräler



Problem

Solution

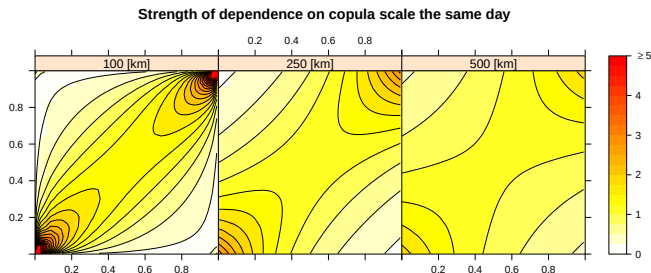
Vine Copulas
Bivariate
Spatio-Temporal
Copulas
Spatio-Temporal Vine
Copulas
Software

Application

Daily Mean PM₁₀

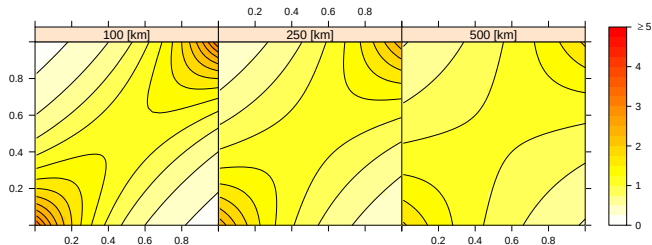
Conclusion & Outlook

References

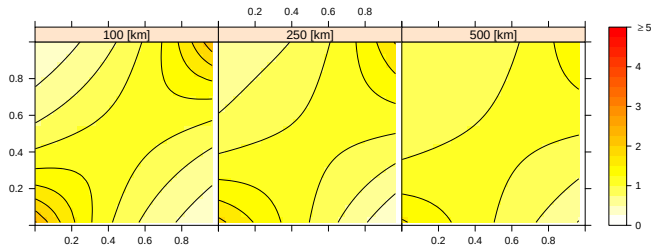


The bivariate spatio-temporal copula IV

Strength of dependence on copula scale one day difference



Strength of dependence on copula scale two days difference



The spcopia
R-package:
Modelling Spatial
and
Spatio-Temporal
Dependence with
Copulas

Benedikt Gräler



Problem

Solution

Vine Copulas
Bivariate
Spatio-Temporal
Copulas
Spatio-Temporal Vine
Copulas
Software

Application

Daily Mean PM₁₀

Conclusion & Outlook

References

The spatio-temporal vine copula I

Build the spatio-temporal neighbourhood and fit the upper vine:

```
> stNeigh <- getStNeighbours(EU_RB_2005, var="rtPM10",
+                             spSize=4, t.lags=-(0:2),
+                             timeSteps=90, min.dist=10)
> stVineFit <- fitCopula(stVineCopula(stConvCop,
+                                     vineCopula(9L)),
+                         stNeigh, method="indeptest")
> stVineFit@loglik
[1] 72417.53
> stVine <- stVineFit@copula
> stVine
```

Spatio-temporal vine copula family with 1 spatio-temporal tree.
Dimension: 10

Problem

Solution

Vine Copulas
Bivariate
Spatio-Temporal
Copulas
Spatio-Temporal Vine
Copulas
Software

Application

Daily Mean PM₁₀

Conclusion & Outlook

References

Predictions can be obtained through

```
> predNeigh <- getStNeighbours(EU_RB_2005, targetGeom,  
+                               "rtPM10", spSize=4, prediction=T)  
  
> stVinePred <- stCopPredict(predNeigh, stVine,  
+                             list(q=qFun), "quantile")
```

Problem

Solution

Vine Copulas
Bivariate
Spatio-Temporal
Copulas
Spatio-Temporal Vine
Copulas
Software

Application

Daily Mean PM₁₀

Conclusion &
Outlook

References

Cross validation results for the daily mean PM_{10} interpolation

	RMSE	bias	MAE
expected value $\hat{Z}_m$	11.20	-0.73	6.95
median $\hat{Z}_{0.5}$	12.08	1.94	6.87
metric cov. kriging	10.69	-0.29	6.28
metric cov. res. kriging	10.67	-0.47	6.16

Table: Cross validation results for the expected value and median estimates following the vine copula approach and two methods from a recent comparison study [GGP12] on spatio-temporal kriging approaches in PM_{10} mapping.

Problem

Solution

Vine Copulas
Bivariate
Spatio-Temporal
Copulas
Spatio-Temporal Vine
Copulas
Software

Application

Daily Mean PM_{10}

Conclusion & Outlook

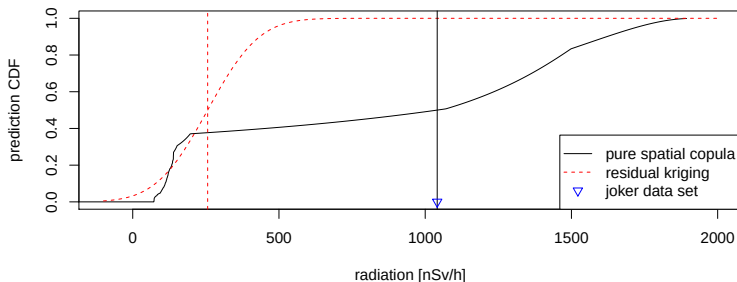
References

Benefits of the spatial/spatio-temporal vine copulas

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

probabilistic advantage flexible uncertainty analysis

The predictive CDFs from the joker data set:



The spcopia
R-package:
Modelling Spatial
and
Spatio-Temporal
Dependence with
Copulas

Benedikt Gräler



Problem

Solution

Vine Copulas
Bivariate
Spatio-Temporal
Copulas
Spatio-Temporal Vine
Copulas
Software

Application

Daily Mean PM₁₀

Conclusion &
Outlook

References

Further extensions

- including covariates
(e.g. altitude, population, ...)
- complex neighbourhoods
(e.g. by spatial direction, ...)
- introduce several spatio-temporal trees in the vine (as in the spatial version, see poster: [Modelling Extremes with the Spatial Vine Copula](#))
- include further copula families
- improve performance

Problem

Solution

Vine Copulas
Bivariate
Spatio-Temporal
Copulas
Spatio-Temporal Vine
Copulas
Software

Application

Daily Mean PM₁₀

Conclusion & Outlook

References

References

-  Kjersti Aas, Claudia Czado, Arnoldo Frigessi, and Henrik Bakken, *Pair-copula constructions of multiple dependence*, Insurance: Mathematics and Economics **44** (2009), 182–198.
-  Tim Bedford and Roger M. Cooke, *Vines – a new graphical model for dependent random variables*, Annals of Statistics **30** (2002), no. 4, 1031–1068.
-  B. Gräler, L. E. Gerharz, and E. Pebesma, *Spatio-temporal analysis and interpolation of PM10 measurements in Europe*, Tech. report, ETC/ACM, 2012.
-  Roger B. Nelsen, *An introduction to copulas*, second ed., Springer Science+Buisness, New York, 2006.

Problem

Solution

Vine Copulas
Bivariate
Spatio-Temporal
Copulas
Spatio-Temporal Vine
Copulas
Software

Application

Daily Mean PM₁₀

Conclusion & Outlook

References