The pair-copula construction for spatial data: a new approach to model spatial dependency

Benedikt Gräler, Edzer Pebesma
Institute for Geoinformatics, University of Muenster

The pair-copula interpolation in brevity

Capturing changing dependence structures

Describing dependence merely by correlation measures or covariance functions reduces the dependence structure of two random variables to a single measure and introduces strong simplifications. The Figure 1 illustrates how the dependence structure of two locations changes over separation distance across the bivariate distribution. Therefore, it can be desirable to model the full multivariate distribution of the observed process.

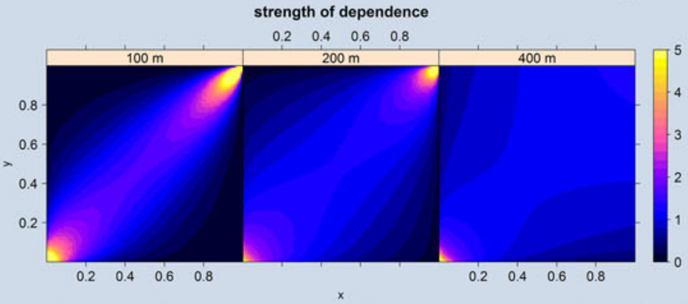

Figure 1: The spatial copula for pairs of locations with different separation distances (in meter). Each tile illustrates the strength of dependence over the full bivariate distribution (the copula's density). The values have been truncated at a level of 5.

Describing dependence with copulas

Copulas enable us to model any n-variate distribution $H(x_1, \dots, x_n)$ by two distinct parts: its n marginal distributions $F_1(x_1), \dots, F_n(x_n)$ and the dependence structure of the margins, its copula $C(u_1, \dots, u_n)$ (see [1]):

$$H(x_1, \dots, x_n) = C\big(F_1(x_1), \dots F_n(x_n)\big).$$

We picked up the idea of the flexible pair-copula construction (PCC; Aas et al. [2]) and adapted it to the spatial framework. Our proposed procedure allows to describe spatial multivariate dependence structures by decomposing them into a set of bivariate copulas.

Pair-copula interpolation vs. ordinary kriging

The application below (Figure 3) demonstrates the potential of the pair-copula interpolation for a local neighbourhood of only 4 neighbours. Cross-validations proved that this first attempt can already compete with ordinary kriging. The root mean square error for the pair-copula interpolation turned out to be close to the one for ordinary kriging, but the bias could be reduced by a factor of ≈ 2.
The full distribution of each estimate provided by the pair-copula interpolation allows for a better uncertainty analysis.

The pair-copula interpolation in more detail

Spatial copula

We developed a *spatial copula* C_h that is a convex combination of copulas. It approaches a copula close to perfect positive dependence for small distances and the product copula (mimicking independence) when the distance tends to the range up to which data are spatially correlated. The convex combination of copulas allows us to mix copula families according to the properties of different distances (as drawn in Figure 1).

Spatial pair-copula over 4 neighbours

We adapt the scheme of the canonical vine (Figure 2) for the *spatial pair-copula*. The copulas of the first tree are replaced with a distance dependent *spatial copula*. The full 5-dimensional copula C_5 is then given by the product of all bivariate copulas.

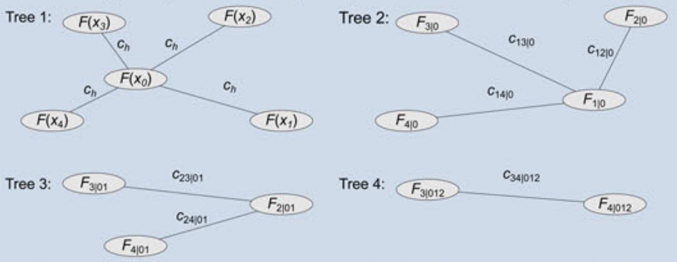

Figure 2: The spatial canonical vine structure. The copulas of the first tree are replaced by a distance dependent spatial copula.

Pair-copula based interpolation

The random process $H(x_0, x_1, \dots, x_4)$ that we model describes the distribution of some variable of interest for a single location and its 4 nearest neighbours. Assuming stationarity allows us to use the same marginal cumulative distribution function F for all locations.
The random variable $Z(s_0)$ at an unobserved location s_0 follows the conditional distribution $H(x_0|x_1, \dots, x_4)$. The density of this distribution can be expressed in terms of a conditional copula. Estimates can then be obtained by calculating the expected value:

$$\hat{x}_0 = \int_0^1 F(u) c_5\big(u|F(x_1), \dots, F(x_4)\big)\, du$$

Meuse river bank

The zinc measurements of the Meuse data set were interpolated using the spatial pair-copula interpolation over 4 neighbours. The margins are modelled by an extreme value distribution. Figure 3 shows the interpolated grid along with the width of the 90% confidence band in measurement and probability scale.

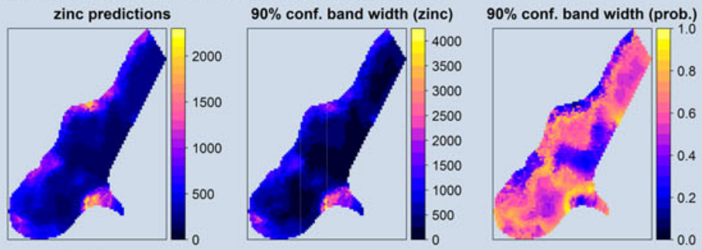

Figure 3: The Meuse river data set interpolated with the proposed pair-copula interpolation (left) and the width of the 90% confidence band in measurement scale (middle) and in probability scale (right).

Comparison of the log-likelihoods of the spatial pair-copula with fits of other 5-dimensional single family copulas shows a considerable higher value for the spatial pair-copula. Providing the wrong distances to the spatial copula in the first tree reduces the log-likelihood as well. Thus, the spatial pair-copula is a good fit within this set of copulas.

Conclusion

The spatial pair-copula strengthens the influence of the spatial information compared to other naïve copula based approaches. It already sufficiently captures the dependence structure of a local neighbourhood to compete with ordinary kriging. The provided conditional distribution for each interpolation location allows for a sophisticated uncertainty analysis of the estimates.

References

[1] Nelsen R. *An Introduction to Copulas.* 2nd ed. New York: Springer 2006.

[2] Aas K, Czado C, Frigessi A, Bakken H. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics* 2009;**44**(2):182-198.

Contact

Benedikt Gräler
Institute for Geoinformatics
Weseler Str. 253
48151 Münster, Germany

ben.graeler@uni-muenster.de
http://ifgi.uni-muenster.de/graeler

ifgi
Institute for Geoinformatics
University of Münster