rtop – an R package for interpolation of data with a non-point support
J. O. Skøien (1), E. J. Pebesma (2), G. Blöschl (3)
(1) Department of Physical Geography, Utrecht University, the Netherlands (2) Institute for geoinformatics, University of Münster (3) Institute for Engineering hydrology and Water Resources Management, Vienna University of Technology

Motivation
It is in ordinary kriging assumed that observations are representative for points or blocks with equal support. This is not the case when we for example look at runoff characteristics (runoff, temperature, floods) or regional health statistics. Catchments are the support in the first case, municipalities or other administrative regions the support in the second case. Several solutions to this problem has been presented, open source, versatile software have been missing.

Background (theory)
- Based on top-kriging method (Skøien et al., 2008) – for prediction of runoff characteristics at locations without observations
- Variogram values between observations and between observations and prediction locations found by integrating a point variogram over a large number of points in each of the catchments (regularization):

\[ \gamma_p(x) = \frac{1}{4} \sum \gamma_h(x,y) \]

- Variogram estimated as cloud variogram or 3-D binned variogram, with areas from each catchment of a pair on the 2nd and 3rd axis
- Point variogram model found by back-calculation (fitting regularized variogram values to sample variogram)
- Kriging equations as normal, can also take measurement uncertainty into account

Example application: Predictions of annual mean
- Annual mean from 387 runoff gauges in Austria
- Cross validation
- Stationarity assumptions can be questioned

Implementation
- Implemented in the statistical environment R (R Development Core Team, 2008)
- Using existing representation of spatial objects in R (Bivand et al., 2008)
- All data and results stored in a single object
- Package created for simple interface with intamap-package (package under development for automatic interpolation through a web-service)
- Format of output similar to gstat-package (Pebesma, 2004)

Usage
> library(rtop)
> rtopObj = createRtopObject(observations, predictionLocations, params)

Example: Point variogram and variogram model

Comparison of sample variogram and variogram model

After fitting a point variogram, sample variogram values can be compared with regularized variogram values from the point variogram for different catchment size classes, or in scatter plots

Interpolated results:
Cross-validation gives correlation between observations and predictions around 0.9, both for point kriging using centre-of-gravity (gstal) and top-kriging. The results are therefore not significantly different from point kriging in this case. Skøien et al. (2006) still found that the method gave results more consistent with expectation in most regions.

Differences from point kriging
Predictions are first of all based on a more sound concept. This can have a large influence on the predictions in some regions, not captured by cross-validation results. A more pronounced difference is the kriging variance. We can either look at the kriging standard deviation or the zscore (residual/kriging standard deviation) as a function of area. In theory, the standard deviation should decrease with increasing prediction area, whereas zscore should be uncorrelated.

Correlations as function of discretization points
- Correlation between observations and predictions stays increase marginally with increasing number of points
- Zscore gets more uncorrelated area with increasing number of points
- Standard deviation is more correlated with area with increasing number of points
- Figure indicates that a minimum of 25 discretization points is recommended

Computation time
- Computation time in the order 5 min (9 points) – 1 hour (100 points) for the examples above
- Some reduction to be expected, most of the time due to numerical integration
- Can be reduced by using geostatistical distance (Gottschalk, 1993) instead of full regularization

Conclusions
- R-package for interpolation of observations with non-point support being developed
- Based on methods from Skøien et al. (2006)
- Planned improvements: more variogram models, more options for variogram fitting, improved graphical output for runoff variables, reduce computation time
- Package will be submitted to CRAN (The Comprehensive R Network), test versions available on request

Acknowledgements
This work was supported by the European Community under the Sixth Framework Programme, in the project “INTAMAP” (Contract N° 033811). The authors acknowledge the INTAMAP team for providing the datasets used in the examples. The authors also wish to thank StatGIS 2009: Surveillance for environmental purposes, June 17-19, 2009, Vienna, Austria. The INTAMAP project is a multi-national project, developed with financial support from the European Commission, and employing open, web-based, data exchange and visualisation tools

References