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)

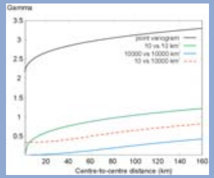
- Based on top-kriging method (Skøien et al, 2006) – for prediction of runoff characteristics at locations without observations
- Variogram values between observations and between observations and prediction locations found by integrating a point variogram over a large number of points in each of the catchments (regularization):

$$\gamma_{12} = 0.5 * Var(z(A_1) - z(A_2)) = \frac{1}{A_1 A_2}\int_{A_1}\int_{A_2}\gamma_p(|\mathbf{x}_1 - \mathbf{x}_2|)d\mathbf{x}_1 d\mathbf{x}_2 -$$

$$0.5 * \left[\frac{1}{A_1^2}\int_{A_1}\int_{A_1}\gamma_p(|\mathbf{x}_1 - \mathbf{x}_2|)d\mathbf{x}_1 d\mathbf{x}_2 + \frac{1}{A_2^2}\int_{A_2}\int_{A_2}\gamma_p(|\mathbf{x}_1 - \mathbf{x}_2|)d\mathbf{x}_1 d\mathbf{x}_2\right]$$

- 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: Point variogram and regularized semivariograms for different combinations of catchment size classes



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)
> # <read data> help functions exist
> rtopObj = createRtopObject(observations, predictionLocations, params)
> rtopObj = rtopFitVariogram(rtopObj)
> rtopObj = rtopKrige(rtopObj)
```

- Observations and predictionLocations are SpatialPolygonsDataFrame (from shapefiles)
- Params includes different options, such as

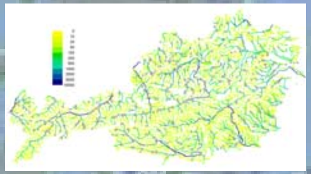
- Cloud/binned variogram	- Variogram model	- Number of discretization points
- Geostatistical distance	- With or without nugget	- Sampling type
- Number of observations	- Limit for kriging weights	- Number of observations for local kriging

INTAMAP

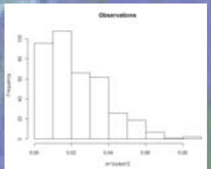
- The INTAMAP project (www.intamap.org) will develop an interoperable framework for real time automatic mapping of critical environmental variables by extending spatial statistical methods and employing open, web-based, data exchange and visualisation tools
- Development case focuses on data from the data base of gamma radiation in Europe – EURDEP – but final software will also include real-time predictions of observations having a support

Example application: Predictions of annual mean

- Annual mean from 387 runoff gauges in Austria
- Cross validation
- Stationarity assumptions can be questioned

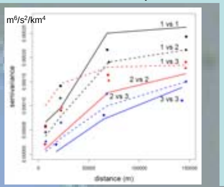


Left: Upslope contributing area (km2)
Right: Histogram of observations ($m^3/s/km^2$)

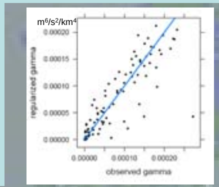


Comparison of sample variogram and variogram model

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

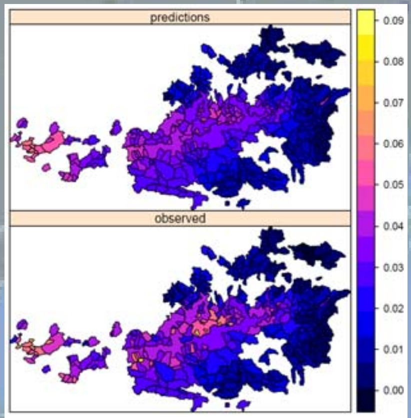


Left: Variograms and cross-variograms for different catchment size classes – observed as lines, regularized as dots and triangles
Right: Scatter plot of observed and regularized semivariogram values



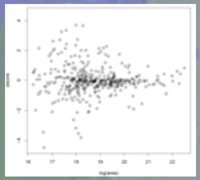
Interpolated results:

Cross-validation gives correlation between observations and predictions around 0.9, both for point kriging using centre-of-gravity (gstat) 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.

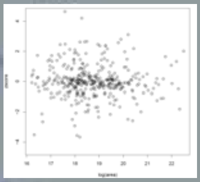


Differences from point kriging

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.

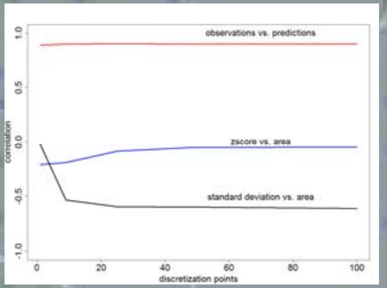


Left: Figures show zscore ((obs-pred)/st.dev) for point kriging and top kriging
Right: Difficult to predict some catchments with high values in central Austria (due to non-stationarity)



Correlations as function of discretization points

- Correlation between observations and predictions only 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

References

Bivand, R. S., E. J. Pebesma, and V. Gómez-Rubio. 2008. *Applied spatial data analysis with R*: Springer.

Gottschalk, L. 1993. Correlation and covariance of runoff. *Stochastic Hydrology and Hydraulics*, *7*, 85-101.

Pebesma, E. J. 2004. Multivariable geostatistics in S: the gstat package. *Computers & Geosciences*, *30*, 683-691.

R Development Core Team. 2008. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Skøien, J. O., R. Merz, and G. Blöschl. 2006. Top-kriging - geostatistics on stream networks. *Hydrology and Earth System Sciences*, *10*, 277-287.

Acknowledgements

This work is funded by the European Commission, under the Sixth Framework Programme, by the Contract N. 033811 with the DG INFSO, action Line IST-2005-2.5.12 ICT for Environmental Risk Management, and from The Austrian Academy of Sciences, project HÖ 18. The views expressed herein are those of the authors and are not necessarily those of the funders.

StatGIS 2009
Surveillance for environmental purposes
Milos, June 17-19, 2009
Contact: Jon Olav Skøien
j.skoien@geo.uu.nl