rtop – an R package for interpolation along the stream network

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

Universiteit Utrecht Faculty of Geosciences

ifgi Institute for Geoinformatics University of Münster

TU Vienna Engineering Hydrology & Water Resources Management

INTAMAP Proposal/Contract no.: 033811 Project start: September 1, 2006 Project end: August 31, 2009 Information Society Technologies

Motivation

- Increased interest in geostatistical methods for variables which has a support
- Examples:
  Regionalisation of runoff variables
  Health statistics
- Support can be spatial and/or temporal
- Methods includes integrals of variogram/covariance functions
- Several solutions to this problem, open source, versatile software still missing

Method

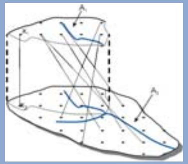
- Based on top-kriging method (Skøien et al, 2006) – for prediction of runoff characteristics at locations without observations
- Variogram values between observations and between observations and prediction locations found by integrating a point variogram over a large number of points in each of the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

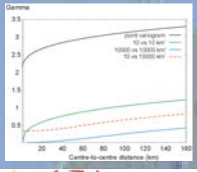
$$\gamma_{12} = 0.5 * Var(z(A_1) - z(A_2)) = \frac{1}{A_1 A_2} \int_{A_1}\int_{A_2} \gamma_p(|\mathbf{x}_1 - \mathbf{x}_2|) d\mathbf{x}_1 d\mathbf{x}_2 -$$

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

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

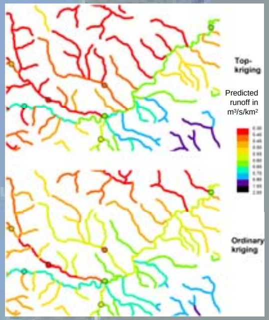
Semivariance found by regularization of a point variogram





Point variogram and semivariograms for different catchment size classes

Example from Skøien et al. (2006) – prediction of 100 year floods (HQ100)



Predictions from Top-kriging consistent with hydrological expectations

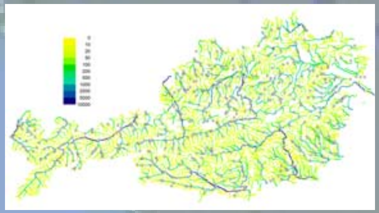
Predictions from Ordinary kriging not consistent with hydrological expectations

Implementation

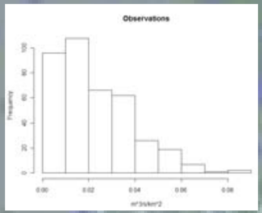
- Method implemented in statistical environment R (R Development Core Team, 2004)
- Taking advantage of existing methods for handling spatial objects (Bivand et al., 2008) and for creating graphical output
- All data and results stored in a single object – for simple interfacing
- Package created for simple interface with intamap-package (package under development for automatic interpolation through a web-service)
- Output similar to output from interpolation in gstat-package (Pebesma, 2004)
- Package will be submitted to CRAN, test versions available on request

Example: Predictions of annual mean

- Annual mean from 387 stations in Austria
- Catchments boundaries for all catchments
- Stationarity assumptions can be questioned

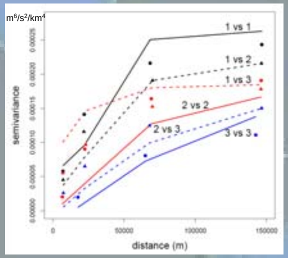


Upslope contributing area in km²

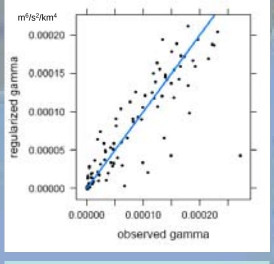


Histogram of observations (m³/s/km²)

Observed and fitted semivariogram values



Variograms and cross-variograms for different catchment size classes – observed as lines, regularized as dots and triangles



Scatter plot of observed and regularized semivariogram values

References

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. 2004. *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.

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

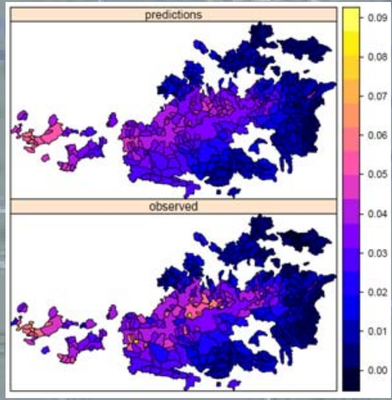
Usage

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

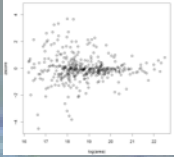
- Observations and predictionLocations are typically SpatialPointsDataFrame (easily created from shapefiles)
- Params includes different options, such as
  - cloud variogram / binned variogram (including parameters for how to make bins)
  - use of geostatistical distance or not
  - variogram model
  - maximum number of observations for local kriging
  - maximum limit for the size of weights for ill-conditioned kriging matrices
- Elements created in earlier analysis can be reused

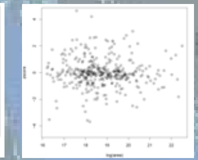
Results:

- Cross-validation gives correlation between observed and predicted around 0.9
- Result not significantly different from point kriging in this case!!
- But: kriging variance dependent on prediction area and zscore (residual/kriging standard deviation) not dependent on area (contrary to point kriging)



Difficult to predict some catchments with high values in central Austria (due to non-stationarity)





Figures show zscore ((obs-pred)/st.dev) for point kriging and top kriging.

Conclusions

- R-package for interpolation of observations with irregular 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
- Package will be submitted to CRAN (The Comprehensive R Network), test versions available on request

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.

EGU General Assembly
Vienna, April 19-24, 2009

Contact: Jon Olav Skøien
j.skoien@geo.uu.nl