Interpolation with irregular support - examining a simplification

J. O. Skøien (1), L. Gottschalk (2), E. Leblois (3)
(1) Department of Physical Geography, Utrecht University, the Netherlands Department of Geosciences, University of Oslo, Norway (3) Cemagref, Lyon, France

Motivation interpolation with a support

- 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

Difficulties with regularization

- Integrations can be slow and lead to numerical instabilities
- Fast and robust methods necessary for real–time interpolation, as developed within the INTAMAP project (www.intamap.org)
- Possible solution: Replacing the integral with an approximation, suggested by Gottschalk (1993)

Approximation

- Suggested by Gottschalk (1993) - replace integration with expectations using Taylor expansion
- The covariance can be expressed through the correlogram:

$$Cov(Z_1, Z_2) = \int_{A_1}\int_{A_2} Cov(|\mathbf{x}_1 - \mathbf{x}_2|)d\mathbf{x}_1 d\mathbf{x}_2 = E\left[Cov(|\,d\,|)\right] = \sigma_x^2 E\left[\rho(d)\right]$$

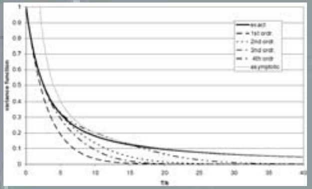
- Where d represents distances between points in the two catchments
- The approximation can similarly be derived for the variogram:

$$\gamma_{12} = 0.5 * Var(z(A_1) - z(A_2)) = \gamma_p(E(|\mathbf{x}_{11} - \mathbf{x}_{22}|)) - 0.5*\left[\gamma_p(E(|\mathbf{x}_{11} - \mathbf{x}_{12}|)) + \gamma_p(E(|\mathbf{x}_{21} - \mathbf{x}_{22}|))\right]$$
$$= \gamma_p(g_{db}) - 0.5\left[\gamma_p(g_{d1}) + \gamma_p(g_{d2})\right]$$

- g_{d1} g_{d2} and g_{db} represent the expected distances between points within the first catchment, the second catchment, and between the two catchments, respectively
- Approximation can generally be referred to as Ghosh approximation from Ghosh (1951)

Example temporal autocorrelation

- Expnential correlation function
- Different orders of Taylor expansion
- T = temporal support relative to correlation length

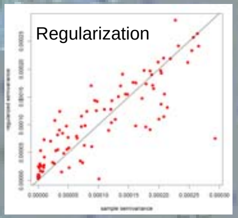


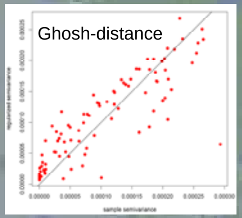
Example: Predictions annual mean flow

- Annual mean flow from 383 stations in Austria
- Top-kriging method (Skøien et al, 2006) used for predictions at locations without observations
- Geostatistical distance used instead of regularization as in original

Comparison variogram values

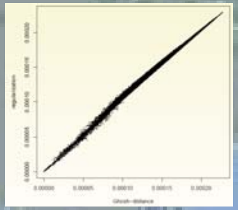
- Sample variogram values (binned) estimated for annual mean
- Figures below show observed versus fitted semivariances for the two methods
- Models are qualitatively similar but give large scatter – probably effect of some violation of stationarity assumptions





Above: Comparison between sample semivariances and fitted semivariances for regularization and Ghosh-distance

Right: Comparison between estimated semivariogram values from same point variogram for regularization and Ghosh-distance



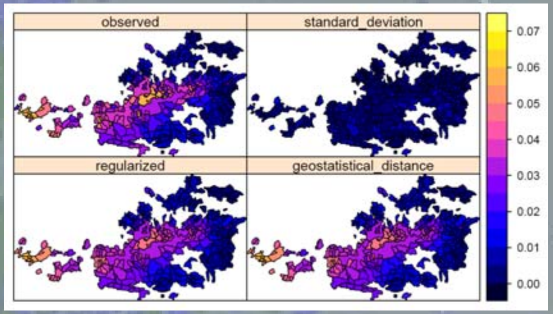
Time consumption

(Just indicative)

Max number of points	Regularization Time (seconds)	Ghosh-distance Time (seconds)
16	19	23
25	24	41
100	135	470
400	1821	7423

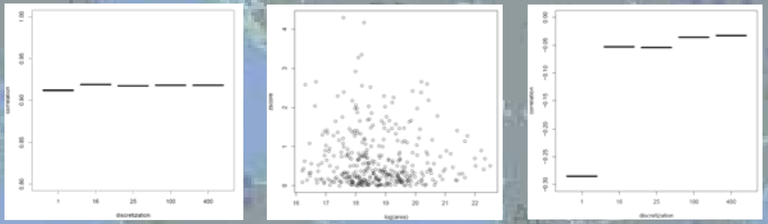
Cross-validation of predictions

- Ghosh approximation does not tend to be more stable than for Top-kriging
- Some very large weights observed
- Below: Comparison of predictions from the two methods, compared with observations and standard deviations
- Units: $m^3/s/km^2$



Effect of number of discretization points

- Number of discretization points limited importance for correlation between observations and predictions (left)
- Correlation between zscore (residual/kriging standard deviation) should ideally be zero
- Strong (negative) correlation between zscore and area for point kriging (middle)
- Correlation decreasing with increasing number of discretization points (right)



Conclusions

- Approximation works in many cases
- Stability of kriging matrix needs to be further checked
- Use of Ghosh-approximation only possibility for real time mapping
- Calculation of ghosh-distances slow, but can be done before real-time mapping takes place

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

References

Gottschalk, L. 1993. Correlation and covariance of runoff. *Stochastic Hydrology and Hydraulics*, *7*, 85-101.
Skøien, J. O., R. Merz, and G. Blöschl. 2006. Top-kriging - geostatistics on stream networks. *Hydrology and Earth System Sciences*, *10*, 277-287.
Ghosh, B. 1951. Random distances within a rectangle and between two rectangles. *Bull. Calcutta Math. Soc.*, *43*, 17-24.

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. The views expressed herein are those of the authors and are not necessarily those of the European Commission.

EGU General Assembly
Vienna, April 19-24, 2009
Contact: Jon Olav Skøien
j.skoien@geo.uu.nl