Interpolating runoff-related variables with rtop

Jon Olav Skøien[1], Günter Blöschl[2], Gregor Laaha[3], Edzer Pebesma[4], Juraj Parajka[5] and Alberto Viglione[6]

Abstract Runoff related variables (runoff, runoff statistics, temperature, concentrations) are often modelled through conceptually or physically based models. The methods are usually data intensive, and require calibration. Geostatistical methods can be regarded as more data driven methods, with a more limited need for data. However, such methods have rarely been applied for hydrological modelling, as the methods were traditionally based on point observations or observations with a regular support. There has recently been a development of geostatistical methods for non-point support, particularly within health modelling (*Goovaerts*, 2006) and hydrology (*Gottschalk, et al.*, 2006; *Skøien, et al.*, 2006). In this paper we will demonstrate an R package (rtop) that implements the methods presented by Skøien et al. [2006], and extended with suggestions from Gottschalk et al. [2006]. Taking advantage of the existing methods in R for manipulating spatially referenced objects (points, lines, polygons, grids), and the extensive possibilities for visualizing the results, rtop makes it considerably easier to apply geostatistical interpolation methods to observations with a non-point support, in comparison to former implementations of the method. Variogram fitting differs strongly from that in traditional geostatistics, but we will present methods for automatic and manual variogram fitting, and for visualizing the model fits. Another feature of the kriging method is that it can handle observation uncertainty, either as a result of measurement uncertainty, or when interpolating statistics, such as annual mean or floods with a certain return period.

[1] European Commission Joint Research Centre, 21027 Ispra, Italy, jon.skoien@jrc.ec.europa.eu

[2] Vienna University of Technology, Institute for Hydraulic and Water Resources Engineering, 1040 Vienna, Austria, bloeschl@hydro.tuwien.ac.at

[3] BOKU Vienna, Institute of Applied Statistics and Computing, 1180 Vienna, Austria, gregor.laaha@boku.ac.at

[4] Institute for Geoinformatics, University of Muenster, D-48151 Muenster, Germany, edzer.pebesma@uni-muenster.de

[5] Vienna University of Technology, Institute for Hydraulic and Water Resources Engineering, 1040 Vienna, Austria, parajka@hydro.tuwien.ac.at

[6] Vienna University of Technology, Institute for Hydraulic and Water Resources Engineering, 1040 Vienna, Austria, viglione@hydro.tuwien.ac.at

Introduction

The change-of-support problem [1] refers to the problem of making geostatistical predictions for an area of a different support than where the observations were made. The necessity for this type of predictions can either be in fields where data are aggregated to administrative units, such as health statistics and forestry data, or where the data are naturally aggregated, such as within hydrology where the runoff at a streamflow gauge is an aggregate of the runoff that has been generated upstream the gauge.

Although there has been developments that try to solve the support problem in geostatistics [2-4], this manuscripts presents what, as far as we know, is the first open source and relatively easily usable software package that applies one of the solutions to the support problem.

rtop is implemented as a package in the statistical environment R [5], which makes it easy to take advantage of existing methods for importing and exporting inputs and outputs and for plotting the results [6].

Theoretical background

The main assumption behind the method is that the variable of interest can be seen as a linear aggregate of a point process in space. This point process is not necessarily observable on the point scale, but can in the case of hydrology be seen as the runoff generation process. Not considering the temporal aggregation in the application in this paper, the runoff at the gauge can then be seen as the sum of the runoff generated upstream the gauge. Skøien and Blöschl [7] used a spatio-temporal aggregation scheme with hydrological routing and a spatio-temporal geostatistical model for interpolation of runoff time series, but although they did not compare their results with an interpolation using only spatial aggregation, it did not appear as if the modeling benefitted from the extra complexity of adding the temporal aggregation. Temporal aggregation should be even less necessary when looking at statistical flows, such as seasonal means or floods/low flows.

The core of the method presented by Skøien et al. [4] is that the observed and modeled semivariance between catchments *i* and *j* can be conceptualized as the regularized point semivariance, using the catchment areas. If A_i and A_j are the areas of the two catchments and γ_p is a point variogram for the continuous process, then the regularized semivariance between the two catchments can be found as:

$$\gamma_{ij} = 0.5 \times \mathrm{Var}(z(A_i) - z(A_j)) = \frac{1}{A_i A_j} \int\limits_{A_i} \int\limits_{A_j} \gamma_p(|\vec{x}_i - \vec{x}_j|) d\vec{x}_i d\vec{x}_j - \quad (1)$$

$$0.5 \times \left[\frac{1}{A_i^2} \int\limits_{A_i} \int\limits_{A_i} \gamma_p(|\vec{x}_i - \vec{x}_j|) d\vec{x}_i d\vec{x} j2 + \frac{1}{A_j^2} \int\limits_{A_j} \int\limits_{A_j} \gamma_p(|\vec{x}_i - \vec{x}_j|) d\vec{x}_i d\vec{x}_j \right].$$

Additionally, the nugget effect is regularized separately as it will diminish for catchment areas of any size with the approach above. Instead, it is assumed that the nugget represents short scale variability that decreases with increasing catchment size in the same way as the variability of the mean of a set of observations decreases with the number of observations. Taking into account that catchments might be overlapping, the nugget effect between two catchments can be found as:

$$C_0(A_1, A_2) = 0.5\left(\frac{C_{0p}}{A_1} + \frac{C_{0p}}{A_2} - \frac{2C_{0p} \cdot \mathrm{Meas}(A_1 \cap A_2)}{A_1 A_2}\right) \quad (2)$$

.

The point variogram is optimized through the iterative optimization procedure SCEUA [8]. The optimization procedure suggests a point variogram that is then regularized either to the catchment sizes of a cloud variogram, or to the catchment sizes and distances of a binned sample variogram. The binned sample variogram is created by grouping the observations of catchment pairs also by their area in addition to their distances. When interpolating, the point variogram is regularized for all combinations of observations and prediction catchments.

The package also includes an option to use a suggestion by Gottschalk [9] to replace the integration in Equation (1) by an integration of the distances between the catchments. This reduces computation time substantially.

The rtop package

The package rtop is a reimplementation and extension of the top-kriging approach, based on the original Fortran version implementing the methods presented in [4]. Some computationally demanding functions have been kept in FORTRAN for faster computation. The implementation links strongly to the sp-package [6], which makes it easy to import and export objects, and to use existing tools for visualizing intermediate and final results. The easiest usage of rtop is through an particular object of rtop-class, but this is only a list of existing object types that makes it possible to call different functions with fewer arguments.

The package also has a simple interface to the intamap-package [10]. This means that it can be used with the interpolation commands in that package, and also easily installed as a part of a geostatistical web service such as intamap[7].

[7] http://www.intamap.org

Example application

In this example we present the results based on data from summer runoff from 138 gauges in the federal county Upper Austria in Austria. In addition to shapefiles of the catchments of the runoff gauges, we also have the shapefiles of 863 catchments were we can predict, and 5775 river segments. This data set is also a part of the package, and can be read with readOGR from the package rgdal [11]. The runoff data has to be normalized with catchment area for use with rtop.

```
rpath = system.file("extdata",package="rtop")
setwd(rpath)
observations = readOGR(".", "obsObero")
predictionLocations = readOGR(".", "predObero")
observations$obs = observations$QSUMMER/observations$AREASQKM
```

An object is then created, including observations, prediction locations, information about the dependent variable and some parameters. In this case the parameters tell rtop to use the simplification from Gottschalk [9] and to use minimum 25 pixels for rasterizing each catchment.

```
rtopObj = createRtopObject(observations, predictionLocations, formulaString =
obs~1,params = list(gDist = TRUE, rresol = 25))
```

The point variogram can be automatically fitted with the function rtopFitVariogram, and the goodness of fit can be visualized in different ways with checkVario.

```
rtopObj = rtopFitVariogram(rtopObj)
rtopObj = checkVario(rtopObj, cloud = TRUE, identify = TRUE, acor = 0.000001)
```

The parameters to checkVario are cloud, which tells the function to show regularized variogram values for each pair, identify, which helps in seeing which catchment pairs make the outliers in the sample variogram, and acor, which modifies the numbers in the plots when the units are of a different magnitude than the data. In the example, the units are meters whereas the catchments are from a few to 2000 square kilometers. Figure 1 shows one of the plots from this function, the point variogram and the regularized variograms for different combinations of catchment areas. The symbol sizes are relative to the number of pairs in each bin.

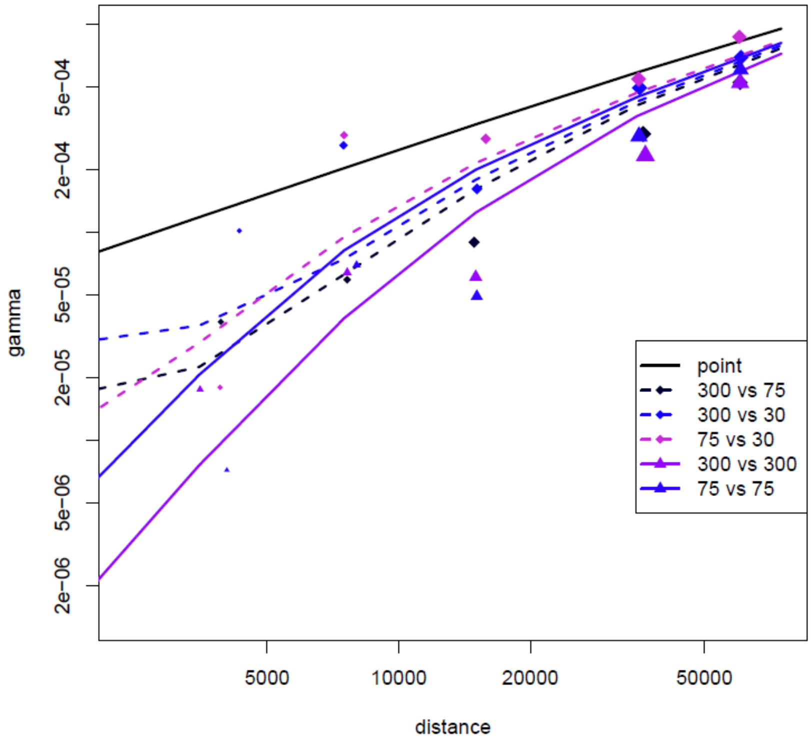

Figure 1 Point variogram and regularized variogram for different combinations of catchment areas

It is then possible to use the point variogram for interpolation or cross-validation (cv = TRUE) with rtopKrige, and then also to look at some summary statistics:

```
rtopObj = rtopKrige(rtopObj,cv=TRUE)
predictions = rtopObj$predictions
cov(predictions$obs, predictions$var1.pred)
summary(predictions)
```

The interpolation has been done for the catchment polygons, whereas it is better to plot along streams. For the example at hand, there are more river segments than predictions. After matching the polygons with the correct stream segment, a function netProp can propagate these values along the network between predictions. Figure 2 shows the specific runoff ($m^3/s/km^2$) predicted for the different catchments.

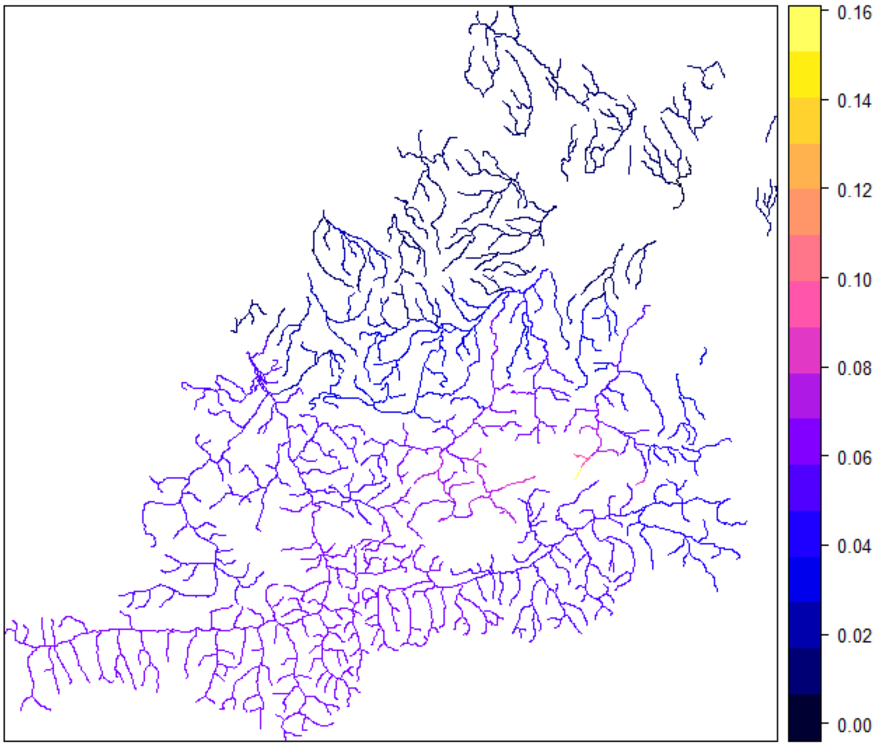

Figure 2 Predictions of streamflow plotted onto the stream network

Conclusions

An R-package for interpolating observations with a non-point support has been developed. This package is relatively easy to use, it is open source, and is developed within the R-environment, which can handle a large range of formats of input and output, and which simplifies creation of graphical output and diagnostics. Being conceptually better based than ordinary point kriging for the same data, it opens up for using geostatistical methods within a field where such methods have only been taken into use in a limited set of applications.

Bibliography

[1] C. A. Gotway and L.J. Young, "Change of support: An inter-discliplinary challenge", in "Geostatistics for environmental applications", P. Renard, H. Demougeot-Renard, and R. Froidevaux, Editors, Springer: Berlin Heidelberg. p. 1-14, 2005.

[2] P. Goovaerts, "Geostatistical analysis of disease data: accounting for spatial support and population density in the isopleth mapping of cancer mortality risk using area-to-point Poisson kriging", International Journal of Health Geographics, **5**: p. art. no. 52, 2006.

[3] L. Gottschalk, I. Krasovskaia, E. Leblois and E. Sauquet, "Mapping mean and variance of runoff in a river basin", Hydrology and Earth System Sciences, **10**: p. 469-484, 2006.

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

[5] R Development Core Team, "R: A language and environment for statistical computing", R Foundation for Statistical Computing: Vienna, Austria, 2012.

[6] R. S. Bivand, E.J. Pebesma and V. Gómez-Rubio, "Applied spatial data analysis with R", Use R, Springer. 378, 2008.

[7] J. O. Skøien and G. Blöschl, "Spatio-temporal Top-kriging of runoff time series", Water Resources Research, **43**: p. W09419, 2007.

[8] Q. Duan, S. Sorooshian and V.K. Gupta, "Effective and Efficient Global Optimization for Conceptual Rainfall-runoff Models", Water Resources Research, **28**(4): p. 1015-1031, 1992.

[9] L. Gottschalk, "Interpolation of runoff applying objective methods", Stochastic Hydrology and Hydraulics, **7**: p. 269-281, 1993.

[10] E. Pebesma et al., "INTAMAP: The design and implementation of an interoperable automated interpolation web service", Computers and Geosciences, **37**(3): p. 343-352, 2011.

[11] T. H. Keitt, R.S. Bivand, E. Pebesma and B. Rowlingson, "rgdal: Bindings for the Geospatial Data Abstraction Library", 2011.