POLISH ASSOCIATION FOR SPATIAL INFORMATION
ANNALS OF GEOMATICS 2008 ❍ VOLUME VI ❍ NUMBER 1

APPLICATION AND VALIDATION OF THE RESIDUAL KRIGING METHOD FOR INTERPOLATION OF THE MONTHLY PRECIPITATION FIELD IN POLAND

ZASTOSOWANIE I WALIDACJA KRIGINGU RESZTOWEGO DO INTERPOLACJI MIESIĘCZNYCH SUM POLA OPADU ATMOSFERYCZNEGO W POLSCE

Maciej Kryza

Department of Meteorology and Climatology, Institute of Geography and Regional Development,
Wrocław University, Poland

Keywords: precipitation, spatial interpolation, residual kriging, Poland
Słowa kluczowe: opady atmosferyczne, interpolacja przestrzenna, kriging resztowy, Polska

Introduction

Spatial interpolation of various elements of climate is an issue of increasing importance (Dobesch et al. 2001). Rainfall, temperature, etc, are spatially continuous phenomena and are key factors for numerical models dealing with many environmental issues, including air pollution transport and deposition, hydrology and hydrogeology. The growing interest on this topic has resulted in the international COST 719 action, focused, among other topics, on GIS applications for spatial interpolation of climatological data (Tveito and Schöner, 2002; Tveito, 2006).

GIS tools provide a wide set of interpolation algorithms, from relatively simple and quick procedures, such as Inverse Distance Weighting (IDW), to more sophisticated and resource consuming tools, such as Kriging. From the climatological perspective, it is useful to distinguish two groups of interpolation procedures: one- and multi-dimensional interpolation methods (Dobesch et al., 2001). In case of one-dimensional schemes, the interpolated phenomena are dependent on the distance between the measuring sites. This is the case for IDW and various kriging algorithms. These methods are widely used in climatology, e.g. for the interpolation of precipitation (Goovaerts, 2000; Stach and Tamulewicz, 2003). However, various elements of climate do not simply change with distance between the measuring sites, but are related to change in a range of other aspects of the environment, with elevation being often the most important (Ninyerola et al., 2000; Ustrnul and Czekierda, 2003). This is where the multi-dimensional interpolation methods can be used, as they make use of additional auxiliary variables (spatially continuous predictors).

This paper presents the application of one of the multi-dimensional interpolation methods, residual kriging (RK), for spatial interpolation of monthly precipitation in Poland, taking October 1975 as an example. The GIS GRASS is used for the interpolation, supported by the R statistical package, both being free Open Source software. The interpolation results are validated and compared with other widely used interpolation algorithms: IDW and ordinary kriging (OK).

Data, methods & software

Measurement data

As this paper deals especially with methodological aspects of the spatial interpolation of precipitation, one month is selected for the detailed study. October 1975 is selected as an example for two reasons:

1. There are a high number of measuring sites (2639), which is of special importance in all (geo)statistical studies. More recent precipitation data are not widely available in Poland, therefore the study is based on these "old" measurements published in the yearbook Opady atmosferyczne 1975.
2. The spatial pattern of precipitation for that month is strongly influenced by both morphological and atmospheric factors.

GIS data and independent variables for multidimensional interpolation

Both atmospheric and morphological factors influencing the spatial pattern of precipitation in October 1975 are considered. Data on atmospheric conditions are taken from the NCEP/NCAR Reanalysis (Kalnay et al., 1996) while terrain elevation is obtained from a SRTM-based digital elevation model (DEM).

The digital elevation model used has 250m x 250m resolution and provides spatially continuous information on terrain elevation. The DEM is used to calculate a number of additional explanatory (independent, auxiliary) variables utilized by the multi-dimensional interpolation algorithms, such as residual kriging (RK). Based on the DEM and GRASS tools, several types of independent variables are calculated and provided as spatially continuous information for the RK interpolation. All DEM derived predictors are calculated with the GRASS tools and they include (Sobik et al., 2001; Kryza, 2008):

1. Mean elevation (ME) of a grid cell neighborhood. The circular shape of the neighborhood is considered with four different diameters: 25, 51, 101 & 301 grid elements (ca 6, 13, 25 and 75 km, respectively).
2. Concavity/convexity factor (CC), calculated as the difference between the DEM (the "real" elevation of a grid cell) and the ME. Grid cells with positive difference are considered as located on a concave form. The CC factors are calculated for all ME diameters.
3. Sheltering index (SI), calculated as the difference between the DEM and the ME calculated for the 90^{o} sector in a given geographical direction. The neighborhood sizes considered are the same as for ME.
4. Terrain roughness (ROU), calculated with the GIS GRASS r.roughness tool (Grohmann, 2006).

Information on elevation is often used in spatial interpolation of various elements of climate (Wiszniewski, 1953; Sobik et al., 2001; Ustrnul and Czekierda, 2003; Kryza et al., 2007). The CC and SI predictors are calculated to describe widely-known processes influencing the spatial pattern of precipitation, for example the rain shadows described by Kostrakiewicz (1975) for the Tatra Mountains in Poland. The terrain roughness can be an important factor in triggering convection processes.

NCEP/NCAR Reanalysis (Kalnay et al., 1997) provides information on atmospheric conditions. These data are used as independent variables in the multi-dimensional interpolation with residual kriging. The NCEP/NCAR data, originally with coarse 2.5^o x 2.5^o resolution, are spatially interpolated to provide a continuous layer on a 1 km x 1 km grid. The NCEP/NCAR data included into the group of potential spatial predictors (i.e. potentially correlated with the measured precipitation) comprise: sea level pressure (SLP); geopotential height (GH); specific humidity (SH); vertical temperature gradient (GRAD); and horizontal wind components (UWND and VWND). The vertical temperature gradient is calculated from the air temperature and geopotential height. The vertical wind component (VWC) is calculated from the UWND, VWND and DEM using the method proposed by Kyriakidis et al. (2001), and included into the group of potential independent variables.

Easting and northing coordinates are also included into the group of potential predictors. The coordinate system used is Polish 1992 CS.

Residual kriging

The spatial field of precipitation is influenced by the circulation regime (meteorological factors) and terrain configuration. If we are able to describe these spatially in a quantitative way (through independent variables described above), then multi-dimensional interpolators can be applied and physical processes that influence the spatial patterns of precipitation can be taken into account in the spatialisation procedure.

Residual kriging (RK) is one of the multi-dimensional interpolation schemes that can be used in climatology. It was successfully applied in Poland for spatial interpolation of air temperature from regional to local scale (Ustrnul and Czekierda, 2003; Szymanowski and Kryza, 2006; Szymanowski et al., 2007; Kryza, 2008). RK is a twofold procedure: firstly, the multiple regression model (MLR) is built, describing deterministic relationships between the independent variables (described above and spatially continuous) and interpolated phenomena, usually provided in climatology as point measurements. Statistically significant independent variables are selected with the stepwise regression procedure. The MLR model is tested both statistically and in terms of physical processes that are behind the selected independent variables. The aim is to build the MLR model that is both statistically and physically meaningful.

In the second step, MLR residuals are spatially interpolated using ordinary kriging. The final map, in this case of monthly precipitation, is calculated by adding the spatial realization of MLR and the interpolated residuals.

Validation

The RK interpolation errors are described with the cross-validation procedure. The “leave one out” scheme is applied here i.e. each time one meteorological station is removed, and the interpolation for the omitted station is performed with the remaining data. The cross-validation error is calculated as a difference between modelled and measured sum of monthly precipitation. The mean errors are described through three statistics (Willmott and Matsuura, 2006):

Mean Absolute Error: $MAE = \frac{1}{N}\sum_{i=1}^{N}\left|E_i\right|$

Root Mean Square Error: $RMSE = \left(\frac{1}{N}\sum_{i=1}^{N}E_i^2\right)^{1/2}$

Mean Error: $BIAS = \frac{1}{N}\sum_{i=1}^{N}E_i$

E_i is the difference between the modelled and measured precipitation for the station *i,* and *N* is the total number of stations. Willmott and Matsuura (2006) suggest that the most appropriate error statistic is MAE. The widely used RMSE generally overestimates the mean information of interpolation error, because it is based on squared cross-validation errors and a small number of large *E* may significantly influence the statistics.

The results of RK interpolation are compared with other widely used algorithms: the inverse distance weighting (IDW) and ordinary kriging (OK; Cressie, 1991). For each method MAE, RMSE and BIAS are calculated with the cross-validation procedure. The t-test is performed to check if the RK cross-validation errors are significantly lower than calculated for other interpolation algorithms.

Software

The spatial interpolation and validation was performed with the GRASS system and R statistical package (GRASS Development Team, 2006; R Development Core Team, 2008) supported with the GSTAT library (Pebesma, 2004), all available as an OpenSource software. These are combined with Linux shell scripts and together made an efficient and flexible set of spatial interpolation tools.

Results

Five independent variables are included in the MLR model of monthly precipitation for October 1975, describing:

1. easting coordinate,
2. sheltering from the north,
3. sheltering from the east,
4. sheltering from the west,
5. vertical temperature gradient.

All independent variables are statistically significant with the p-value <0.05. The adjusted R^2 is close to 0.63. The signs of each of the regression coefficients can be explained in terms of physical processes responsible for the spatial pattern of the precipitation in this month. The relation between measured precipitation and sheltering indexes is related to the prevailing wind direction, as described by Kostrakiewicz (1975), but detailed description of the synoptic situation of the month selected is beyond the scope of this study. Positive correlation found for SI for north and west direction describes processes related to the ascending air, whereas the descending air flow is described by the negative beta coefficient for the east SI.

The beta coefficient for GRAD is greater then zero, therefore higher precipitation is modelled over the areas with greater atmospheric instability. The negative relation with the easting coordinate may be linked with the increasing distance from the Atlantic Ocean.

It should be noticed that the elevation a.s.l. is not statistically significant in this case. This is of special importance, as the elevation-based auxiliary variables are often used in multidimensional interpolation (Wiszniewski, 1953; Kozyra, 2006).

The final map of monthly precipitation is presented in Figure 1. The role of atmospheric factors is clearly visible over the central part of Poland, where monthly precipitation for this lowland area is at a similar level as for the mountains in the south. The areas of high rainfall sums have generally larger cross validation errors.

Table. Cross-validation errors

	RK	OK	IDW
MAE	8.30	9.17	9.18
RMSE	10.98	12.14	12.22
BIAS	0.00	0.11	0.08

The MAE, RMSE and BIAS error statistics for RK are lower than for the other methods tested (Table). The differences are statistically significant at p-value < 0.05. The largest errors are calculated for the simple IDW, but the results achieved for OK are not much different. The positive BIAS for OK and IDW tend to overestimate the precipitation height.

If the cross-validation errors are analyzed in the vertical profile, there is a positive correlation with elevation for all three interpolation algorithms (Fig. 2) and, in general, the CV errors for RK are smaller than for IDW and OK. It should be mentioned that the MAE errors for the areas above 800 m a.s.l. are based on data from a relatively small number of stations (less then five), and therefore might be considered as statistically insignificant. The IDW method performs better than RK and OK at elevation of 900 m a.s.l. but this doesn't influence significantly the overall poor MAE and RMSE statistics.

A visual comparison of the interpolation results was performed for a selected mountainous region in south Poland (Fig. 3; location of the study area – see Fig. 1). The differences are rather large, and the RK may be considered as the optimal interpolation scheme, taking into account the general climatological features of the area. Visually most dubious is the IDW method due to the strong bulls-eye effects.

Summary & discussion

The quantitative and visual validation and comparison of the interpolation results of monthly precipitation for October 1975, calculated with three different algorithms, have shown that the residual kriging method gives the smallest cross-validation errors and visually the most reliable results from the climatological point of view. The cross-validation errors calculated for OK and IDW are significantly higher, and the interpolated spatial pattern of the precipitation is dubious or unacceptable, in case of IDW, because of the strong bulls-eye effect.

The main advantage of the RK is the use of the auxiliary variables on the first step of the interpolation scheme. This is also the field of future improvements, as the set of the potentially significant predictors can and should be widened to reflect a greater number of atmospheric processes influencing the spatial pattern of precipitation.

The GRASS software, supported with R statistical package, proved to be very efficient and flexible tools for both spatial interpolation and validation of the results. The flexibility can be improved considerably if GRASS and R are combined with the Linux shell scripts. All these packages are available as free OpenSource software, therefore being a good option for researchers & users with limited funding.

References

Cressie NAC 1991: Statistics for spatial data. John Wiley&Sons.

Dobesch H., Tveito OE., Bessemoulin P., 2001: Geographic Information Systems in Climatological Application. DNMI-Report 13/01 KLIMA.

Goovaerts P., 2000: Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *J. Hydrology* 228(1-2): 113-129.

GRASS Development Team, 2006: Geographic Resources Analysis Support System (GRASS) Software. ITC-irst, Trento, Italy. http://grass.itc.it.

Grohmann CH. 2006: r.roughness – a new tool for morphometric analysis in GRASS. GRASS/OSGeo-News 4: 17-19.

Kalnay E., Kanamitsu M., Kistler R., Collins W., Deaven D., Gandin L., Iredell M., Saha S., White G., Woollen J., Zhu Y., Leetmaa A., Reynolds B., Chelliah M., Ebisuzaki W., Higgins W., Janowiak J., Mo KC., Ropelewski C., Wang J., Jenne R., Joseph D., 1996: The NCEP/NCAR 40-Year Reanalysis Project. Bulletin of the American Meteorological Society 77: 437-471.

Kostrakiewicz L., 1975: Modele matematyczne zależności średnich sum opadów atmosferycznych w Karpatach Polskich od hipsometrii, ekspozycji i rzeźby terenu oraz długości geograficznej. *Przegląd Geofizyczny* XLIX: 53-65.

Kozyra J., 2006: Zastosowanie Modelu Agroklimatu Polski (MAP) do tworzenia numerycznych map opadów atmosferycznych w Polsce. [in:] Migała K., Ropuszyński P. (eds.), Współczesna meteorologia i klimatologia w geografii i ochronie środowiska, PTG Wrocław: 69-76.

Kryza M., 2008: Modelowanie pola opadu atmosferycznego jako uwarunkowania procesu depozycji zanieczyszczeń. Praca doktorska, Uniwersytet Wrocławski.

Kryza M., Szymanowski M., Wieczorek M., 2007: Wybrane metody interpolacji w modelowaniu map ekstremalnej temperatury powietrza (na przykładzie Polski południowo-zachodniej). *Przegląd Geofizyczny* LII(1): 61-82.

Kyriakidis PC., Kim J., Miller NL., 2001: Geostatistical mapping of precipitation from rain gauge data using atmospheric and terrain characteristics. *J. of Appl. Met.* 40: 1855-1877.

Ninyerola M., Pons X., Roure JM., 2000: A methodological approach of climatological modeling of air temperature and precipitation through GIS techniques. *Int. J. Clim.* 20: 1823-1841.

Opady atmosferyczne 1961-1980, IMGW, Wyd. Komunik. i Łączności

Pebesma EJ., 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. R Foundation for Statistical Computing, www.R-project.org.

Sobik M., Netzel P., Quiel F., 2001: Zastosowanie modelu rastrowego do określania pola rocznej sumy opadów atmosferycznych na Dolnym Śląsku. *Rocznik Fizyczno-Geograficzny*, Uniwersytet Gdański 4: 27-34.

Stach A., Tamulewicz J., 2003: Wstępna ocena przydatności wybranych algorytmów przestrzennej estymacji miesięcznych i rocznych sum opadów na obszarze Polski. [In:] Kostrzewski A., Szpikowski J. (eds.), Funkcjonowanie geoekosystemów zlewni rzecznych, UAM Poznań: 87-111.

Szymanowski M., Kryza M., 2006: Evaluation of selected spatialisation techniques in urban heat island estimation. 6th International Conference on Urban Climate Göteborg, Sweden, June 12th-16th 2006: 435-438.

Szymanowski M., Kryza M., Smaza M., 2007: A GIS approach to spatialize selected climatological parameters for wine-growing in Lower Silesia, Poland. Proceedings of the Bioclimatology and Natural Hazards Conference, 17-20.09.2007, Polana, Słowacja.

Tveito OE., Schöner W., 2002: Applications of spatial interpolation of climatological and meteorological elements by the use of geographical information systems (GIS). KLIMA Report 28/02.

Tveito OE. (ed.), 2006: Spatialisation of the climatological and meteorological information by the support of GIS. COST 719 Final Report WG2 Spatialisation (Final draft version 26.07.2006).

Ustrnul Z., Czekierda D., 2003: Zróżnicowanie przestrzenne warunków termicznych powietrza Polski z wykorzystaniem GIS. IMGW seria: Atlasy i Monografie.

Willmott CJ., Matsuura K., 2006: On the use of dimensioned measures of error to evaluate the performance of spatial interpolators. *Int. J. of GIS* 20(1): 89-102.

Wiszniewski W. (ed.), 1953: Atlas opadów atmosferycznych w Polsce. PIHM Warszawa.

Abstract

Residual kriging (RK) is applied to calculate spatial patterns of monthly precipitation in Poland. RK is one of the multi-dimensional interpolation schemes being used for spatial interpolation of various climatological data. The algorithm is supported by additional explanatory variables, which reflect physical processes responsible for spatial pattern of the interpolated phenomena. Here, the algorithm is implemented in the OpenSource GIS GRASS, supported by the R statistical package. These tools are found to be very flexible and efficient for this purpose.

The RK results are compared with other interpolation procedures (Inverse Distance Method and Ordinary Kriging). Both visually and in terms of cross-validation errors, RK performs better than the other interpolation algorithms tested. The general statistics, Root Mean Square Errors and Mean Absolute Errors, are lower for RK than for the other methods compared, and the differences are statistically significant. The cross-validation errors are strongly correlated with the areas of high precipitation. There is also an increase of cross-validation error with the elevation a.s.l., but the increase is smaller for the RK than for the other interpolation algorithms used.

Streszczenie

W pracy zastosowano metodę krigingu resztowego (Residual Kriging, RK) do przestrzennej interpolacji miesięcznych sum opadu atmosferycznego w Polsce. Kriging resztowy jest jednym z algorytmów interpolacji wielowymiarowej, który znajduje szerokie zastosowanie w klimatologii. Wynika to z faktu, że RK uwzględnia dodatkowe zmienne niezależne (objaśniające), odpowiedzialne za przestrzenny opis procesów fizycznych kształtujących interpolowane zjawisko. W tej pracy algorytm zaimplementowany został w systemie GIS GRASS, współpracującym z pakietem statystycznym R. Obydwa środowiska działają na licencji OpenSource i okazały się być bardzo wydajnymi narzędziami do realizacji prezentowanego zadania.

Wyniki interpolacji RK zostały porównane z otrzymanymi z zastosowaniem dwóch innych algorytmów, metody odwrotnych odległości (IDW) oraz krigingu zwykłego (OK). Porównanie zostało wykonane zarówno jakościowo (wizualnie), jak i ilościowo przez zastosowanie oceny krzyżowej. W obu przypadkach wyniki uzyskane RK są lepsze niż dla IDW i OK. Ogólne statystyki błędu oceny krzyżowej, RMSE i MAE, policzone dla RK, są niższe niż uzyskane dla IDW i OK, a różnice są istotne statystycznie. Wysokie błędy interpolacji są skorelowane przestrzennie z obszarami z wysokimi sumami opadów. Stwierdzono także wzrost błędów oceny krzyżowej z wysokością, który dla RK jest mniejszy niż dla IDW i OK.

dr Maciej Kryza
kryzam@meteo.uni.wroc.pl
phone: +48 71 348 54 41
www.meteo.uni.wroc.pl

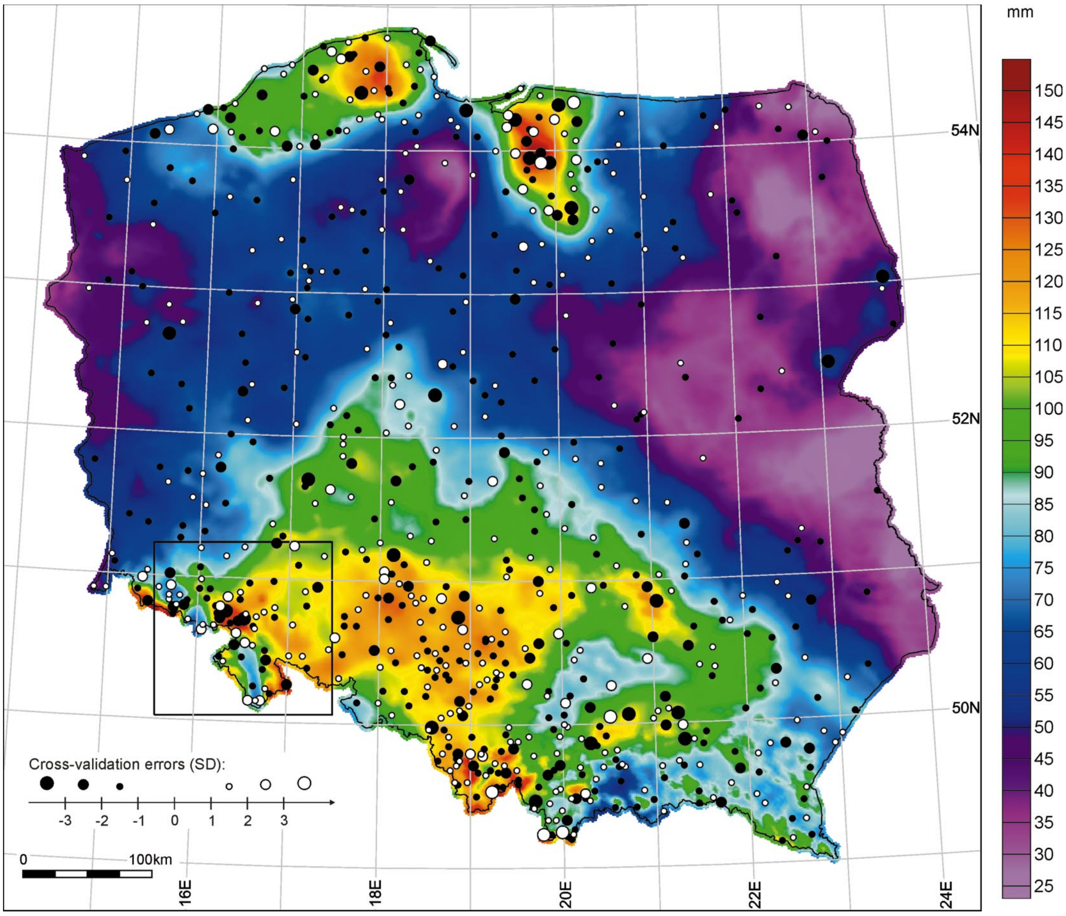

Figure 1. Monthly precipitation in October 1975 interpolated with RK and cross-validation errors (SD-standard deviation; black rectangle – area of the detailed analysis)

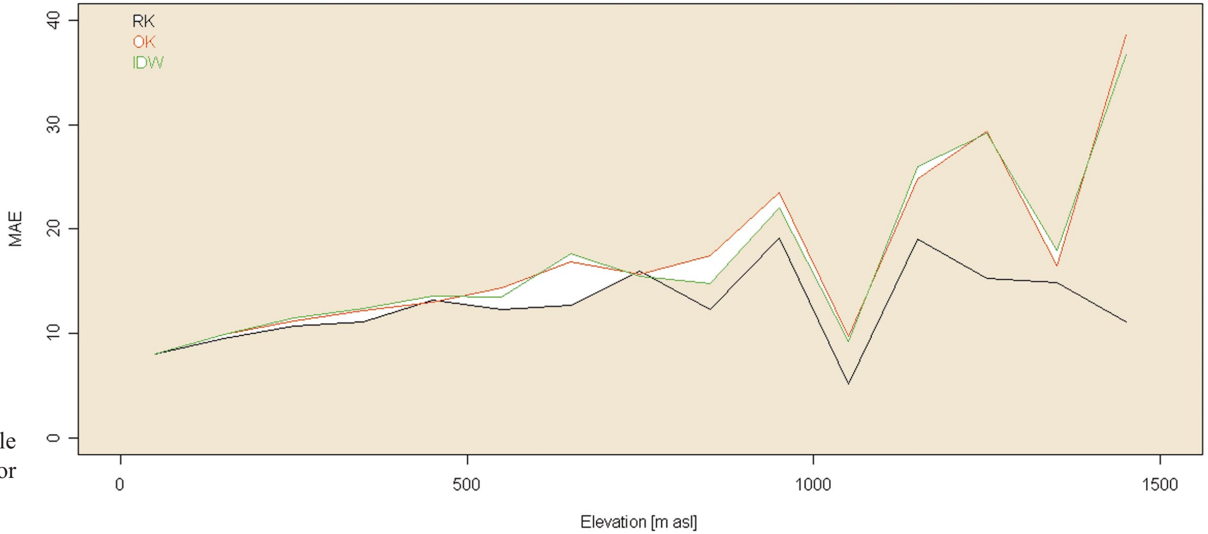

Figure 2. Vertical profile of the cross-validation error

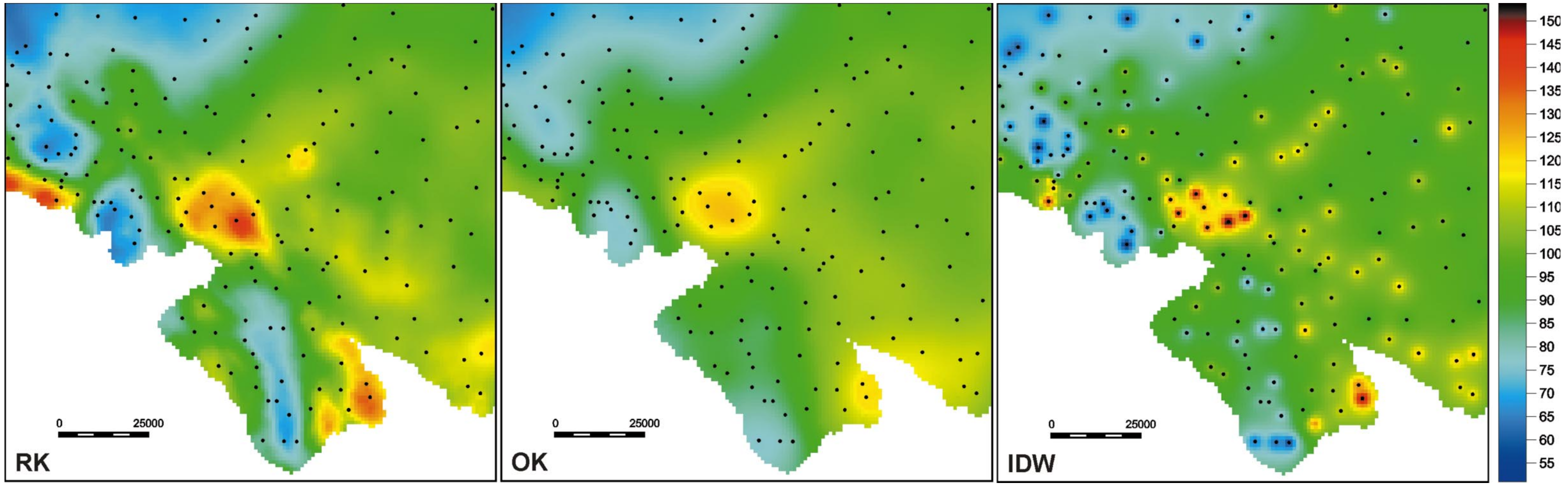

Figure 3. Interpolation results for RK, OK and IDW (black dots – precipitation measuring sites)