ABOUT THE PROBLEM OF DIGITAL PRECIPITATIONS MAPPING USING (GEO)STATISTICAL METHODS IN GIS

C.V. Patriche[1], L. Sfîcă[2], B. Roşca[3]

ABSTRACT

Our study reveals some problems related to the digital mapping of mean annual precipitations using (geo)statistical methods in GIS environment. The applications are carried out for a 4950 km^2 region situated in Vrancea County, using a sample of 34 rain gauges. We first address the data uncertainty issue looking for georeference errors and data errors. As a result we chose to eliminate 2 rain gauges from our analysis significantly evading the general spatial precipitations pattern probably due to the shorter data recording intervals. We show how easily such outliers can mislead us by inducing a false precipitations – latitude correlation. We then proceed by deriving mean annual precipitation spatial models using classical statistical approaches (ordinary kriging, cokriging) and a more elaborated approach (residual kriging). Comparison of the results proves the superiority of the latter. Still, the uncertainty of the output has to be considered, especially when it comes to the extrapolation of the residual kriging model outside the calibration area.

*

1. INTRODUCTION

The atmospheric precipitations constitute a very important climatic element, being the input data for many models of different nature, such as (agro)climatic, hydrological, biological, soil models etc. From all climatic parameters, the precipitations are probably the most difficult to model, due to their main dependency on air masses dynamics. As in the case of other climatic parameters, the space and time scales influence greatly the choice of the spatialisation method, the accuracy of the output etc. (*Patriche C.V.*, 2007). The most difficult to model are the momentarily values (e.g. single rainfall event, daily precipitations) as they do not show an important dependency on the topographic or other quantifiable terrain characteristics. At such scales, simple local interpolators like ordinary kriging, cokriging, IDW, are sufficient for deriving fairly good spatial models. Mean values, such as mean monthly or annual values are generally more predictable. Still, we often find out that, except for the local altitude, other terrain characteristics do not explain much of the precipitations variance.

Many spatialisation methods can be applied for deriving spatial models of precipitations (*Patriche, C.V.*, 2005). A good synthesis about the use of statistical spatialisation methods for meteo-climatic variables is given by *Dobesch, H., Dumolard, P., Dyras, I.* (editors, 2007) and about the use of geostatistical (kriging) methods by *Hengl, T.* (2007). We mentioned already, local interpolators such as ordinary kriging, cokriging, IDW, which are more suitable for momentarily values. In addition, we may mention the

[1] *Romanian Academy, Department of Iasi, Geography Group*
[2] *"Al. I. Cuza" University of Iasi*
[3] *"Stejarul" Center for Biological Research, Piatra Neamt*

universal kriging, which computes first a polynomial trend surface, and especially the widely used residual kriging, also known as detrended kriging or regression kriging. The last method uses a combination of regression and ordinary kriging to derive both terrain dependable (regional) precipitations characteristics and local precipitations characteristics.

2. STUDY REGION

Our study focuses on problems related to spatialisation of mean annual precipitation values at regional scale. The study region is situated mainly in Vrancea County and covers a surface of 4950 km^2, comprising a relief of plains, hills and mountains, with altitudes generally increasing westwards, from 4m to 1770m (fig. 1). We used precipitation data from a sample of 34 rain gauges and a Digital Elevation Model with a resolution of 30m.

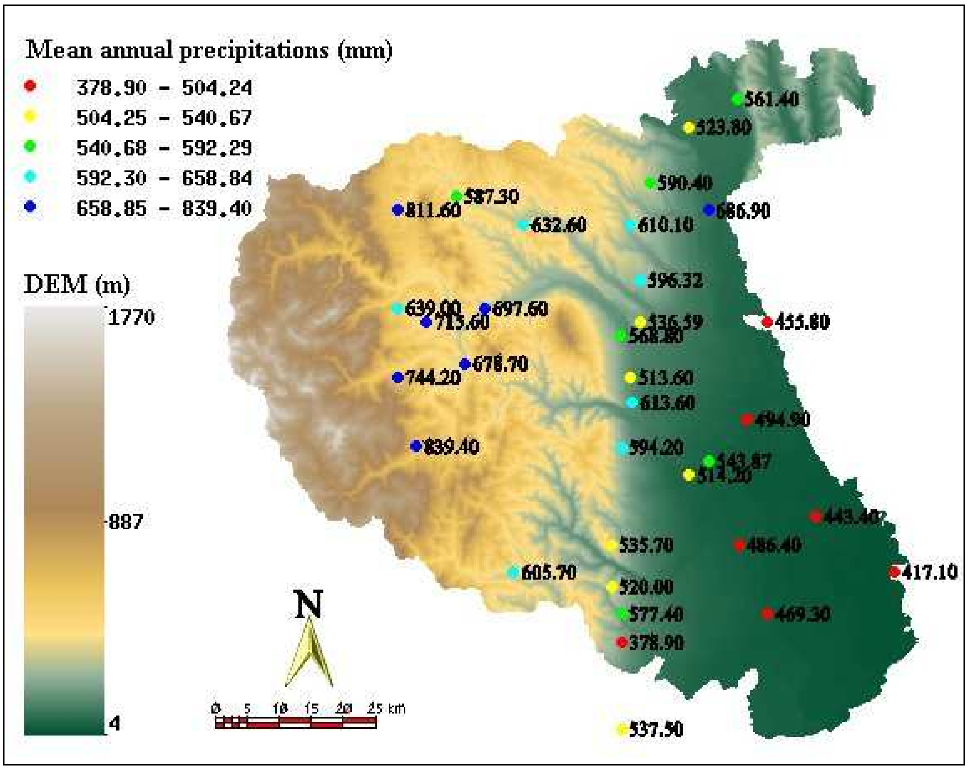


Fig. 1 ***Using the preliminary visualization of point precipitation data to search for possible outliers and precipitations spatial patterns***

3. INPUT DATA UNCERTAINTY ISSUES

Before starting our statistical analysis, a *visual inspection* of mapped rain gauge precipitation data may be useful for identifying possible outliers and precipitations spatial patterns. Figure 1 shows the mapped point precipitation data classified into intervals including an equal number of points. We may easily see the general pattern of precipitations increase from east to west caused by the increase of the altitude in the same direction. We may also notice the *bias* of the rain gauges network caused by its preferential location at lower altitudes, along the valleys and at the contact between plains and hills (piedmont area). The mountainous region, situated in the west, is practically uncovered by rain gauges, meaning that our models will have to be extrapolated here and we shall have to decide if the extrapolation is realistic or not.

The visual inspection of mapped rain gauge precipitation data also points out the presence of 2 possible outliers, Pufesti (686.9mm) and Slobozia Bradului (378.9mm). We shall refer to them later.

Let as first analyze another source of uncertainty, which is often overlooked: *the georeference errors*. Georeference errors refer to errors of the X, Y, Z coordinates. Misplacements of station / rain gauges points on the map may induce significant errors, especially in highly fragmented terrain, when predictors' values are extracted from raster layers or when local interpolators such as kriging are used for spatial modeling. The former will lead to wrong predictors' values and therefore inaccurate regression models, while the latter will generate locally displaced precipitation fields.

Correlation between the stations / rain gauges altitudes and the respective DEM altitudes may be used for identifying possible georeference errors or errors in recording the stations / rain gauges altitudes (fig. 2). The correlation should be very good, although not perfect for several reasons: the DEM generalizes the altitude information according to its resolution; the stations / rain gauges latitude and longitude values are generally given in degrees and minutes. Now supposing that the seconds are rounded up or down to the closest minute, it actually means that we may have a coordinate error of up to 30 seconds, meaning about 900m for latitude and 600m for longitude. These errors double if no coordinate rounding was performed and the seconds were just disregarded.

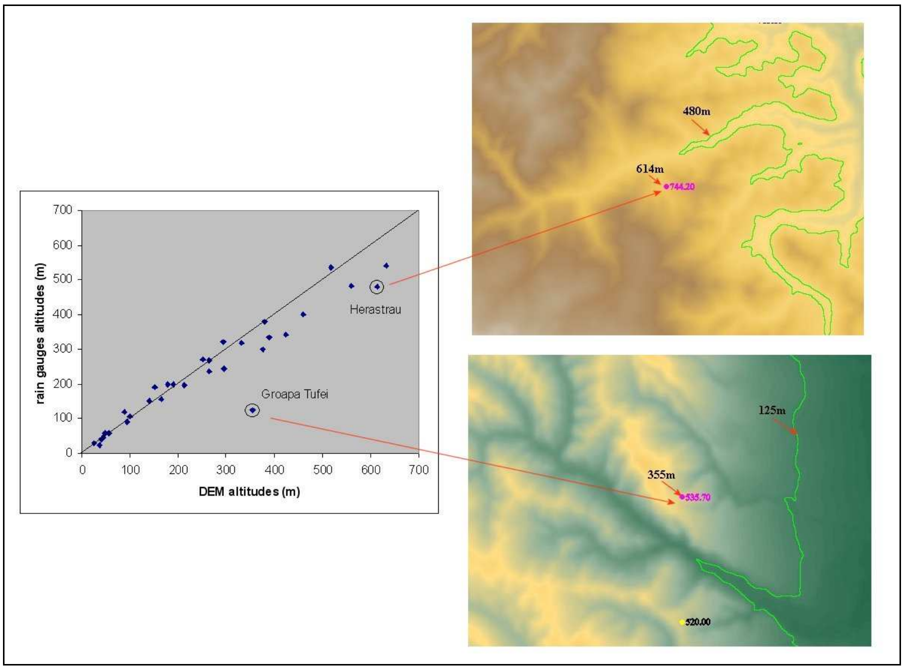


Fig. 2 ***Correlation between the rain gauges altitudes and the respective DEM altitudes as a method to identify possible georeference errors***

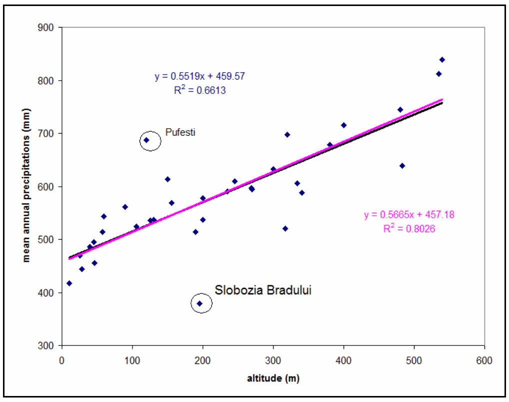
In our situation, we notice 2 points situated outside the correlation cloud indicating possible georeference errors: Groapa Tufei and Herastrau (fig. 2). The position error is very obvious for Groapa Tufei, in which case the recorded rain gauge altitude is 125m, while the DEM altitude for this particular location is 355m. We can see how far away is the 125m altitude isoline along which the rain gauge should be located. There are 2 possible explanations for this error: either the horizontal coordinates of Groapa Tufei are wrong, or the recorded altitude is incorrect. What is the potential negative impact of such a georeference error on spatial statistical models of precipitations? If the real altitude of Groapa Tufei is 125m, so the recorded altitude is correct, but the horizontal coordinates are wrong, then this point may be used for regression analysis, provided that the neither the DEM altitude values or other derived predictors' values are used for models computation. In a geostatistical approach (ordinary kriging, residual kriging etc.) it is not advisable to include such misplaced points because they will misplace, in their turn, the precipitation values. Still, if the value of a misplaced point is similar to those of the neighboring points, the error induced by the georeference error may be small enough and the respective point may be kept. This is also the case of our 2 georeference errors displayed in figure 2.

Let us refer now to the possible *data errors*. As we mentioned before, there are 2 points situated significantly outside the altitude – precipitations correlation cloud, namely Pufesti (686.9mm) and Slobozia Bradului (378.9mm), therefore indicating the presence of 2 possible outliers (fig. 3). In the case of Pufesti rain gauge, the mean annual precipitation regime is characterized by a secondary maximum in August. Taking into account that all other rain gauges display a single maximum in June, we are inclined to believe that either the August data is incorrect or the Pufesti data represent a shorter time frame, corresponding to a more humid period. On the other hand, the mean annual value recorded at Slobozia Bradului rain gauge is obviously too small for the climatic conditions of our region. Because the monthly values display a normal annual distribution, we are inclined to believe, as before, that the data correspond to a shorter time frame from a drier period.

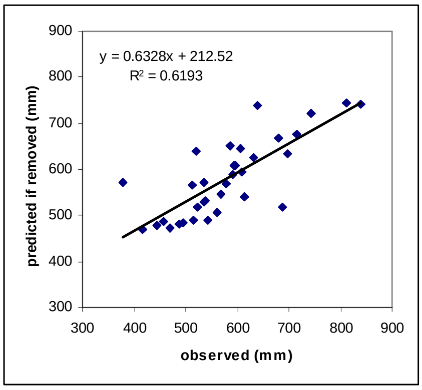
But how do these points affect our spatial precipitation model? Is it necessary to remove them from analysis? Figure 3 shows the influence on the altitude regression model. We notice that even though these 2 points are associated with the highest residuals, the difference between the actual and the deleted residuals (jackknife error) is small (fig. 3c), meaning that their removal from analysis does not significantly change the altitude regression model. This is happening because the points are situated on opposite sides as compared to the regression line (fig. 3a) and therefore have opposite effects, balancing the regression line. Their removal increases the correlation coefficient but does not significantly change the direction of the regression line, meaning that the regression equations are very similar with or without these points. This can also be grasped if one notices that the altitude – precipitation correlation coefficient (0.66) is quite similar with the cross-validation correlation coefficient (0,62), meaning that the one by one removal of all sample points does not significantly change the altitude – precipitations relationship.

What about the effects on other predictors? We must mention that, apart from altitude, we also used latitude and longitude as predictors and at first we obtained a good regression model using both altitude and latitude. Looking further into details, we noticed that the latitude – precipitations correlation is a false correlation, induced by the presence of the 2 outliers (fig. 4), one with a higher precipitation value situated in the northern part of our region (Pufesti), the other one with a lower precipitation value situated in the South (Slobozia Bradului). If one eliminates these 2 points, the latitudinal correlation cannot be depicted any more.

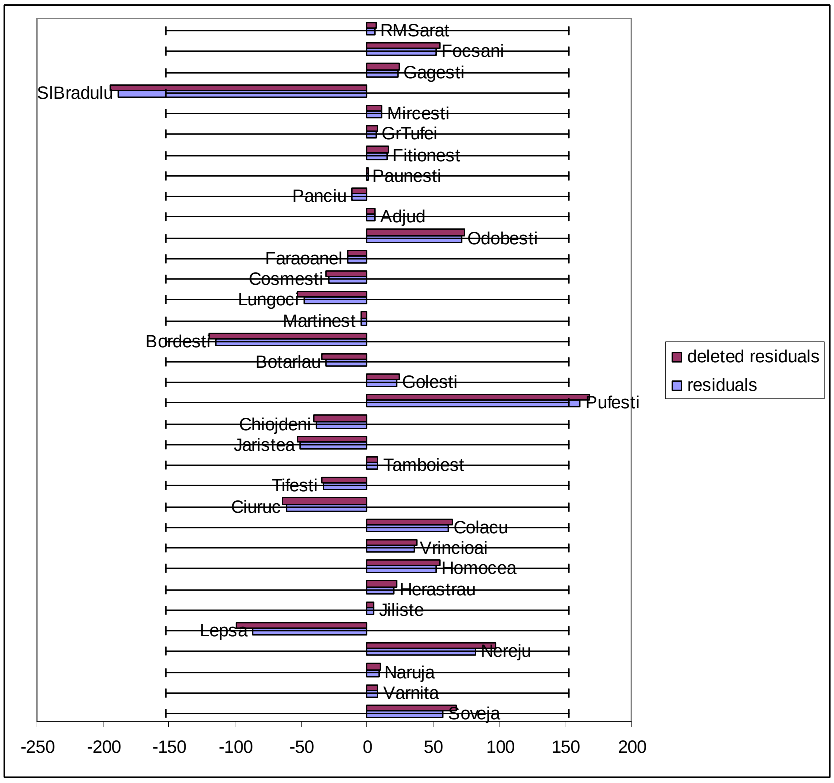
For this reason and because of our intention of using also kriging for spatialisation, in which case the great residual values of the 2 suspect points would be represented on the map, we decided to eliminate them from analysis.



a



b



c

Fig. 3 ***The altitude – mean annual precipitations relationship (a) and comparison between actual and deleted residuals (c) showing the presence of 2 possible outliers; cross-validation of the altitude model using all stations (b)***

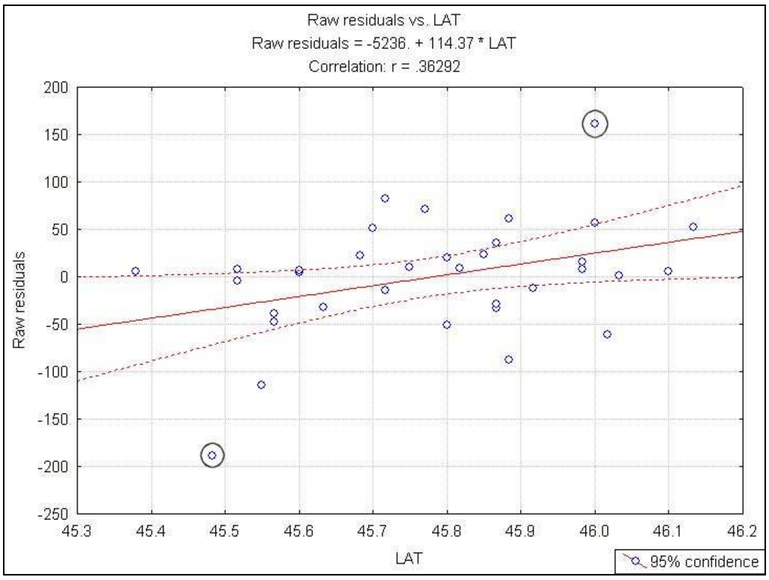


Fig. 4 ***An unwanted effect of outliers: the false correlation precipitations – latitude***

4. SPATIALISATION OF MEAN ANNUAL PRECIPITATIONS. RESIDUAL KRIGING VS. ORDINARY KRIGING (COKRIGING)

Resuming the data uncertainty issues, we recall that we found 2 important georeference errors corresponding to Groapa Tufei and Herastrau rain gauges, but we decided that we may still keep these points because their values are very similar to those of the neighboring points, so the errors induced by their presence are negligible. On the other hand, we found 2 suspect data points (Pufesti and Slobozia Bradului), showing values very different from those of the neighboring points, due either to data errors or, more likely, to shorter precipitation recording intervals. Because their presence may significantly affect a geostatistical model and because they induce a false latitudinal correlation, we have decided to eliminate them from our analysis.

The spatialisation of the mean annual precipitations was carried out by means of 3 different methods, namely ordinary kriging, cokriging and residual kriging, in order to compare their performances. Because the cokriging spatialisation, with altitude as co-variable, proved to be very similar to the ordinary kriging spatialisation, we shall not refer to it as it follows, because the conclusions regarding the ordinary kriging output apply also to the cokriging approach.

Figure 5 shows *the ordinary kriging* spatial model of mean annual precipitations. We may notice the smoothness of the precipitation field with values gradually increasing from east to west, following the general increase of the terrain altitude.

The *residual kriging* approach (fig. 6) combines regression and ordinary kriging methods to produce the final map. Regression analysis is used to link the predictand values to the terrain characteristics, while the ordinary kriging models the regression residuals. Finally, the two spatial models are added up to produce the final map.

In our case, we used the altitude regression model achieved after the elimination of the 2 suspect points referred to previously (fig. 7). The explained variance is 80% and the estimated mean annual precipitation vertical gradient is about 56mm/100m. The regression line of the correlation between the observed and the predicted values is very close to the main diagonal, along which it should be situated in an ideal situation.

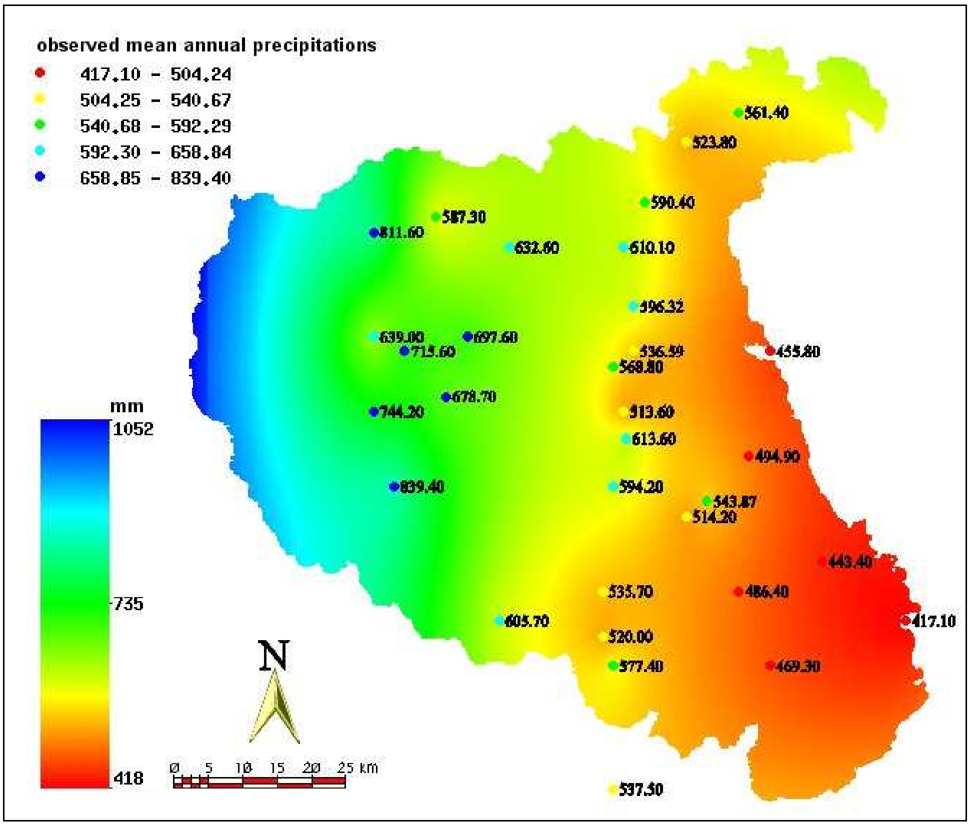


Fig. 5 ***Ordinary kriging of observed data***

One of the problems related to regression models is that of model's *extrapolation*. In our case, the westernmost part of the region is uncovered by rain gauges, meaning that we shall have to extrapolate our regression model there if we want to estimate the mean annual precipitation values for this part as well. Performing the extrapolation up to 1770m of altitude, we estimate precipitation values of up to 1463mm. Such estimated values are, in our opinion, unrealistic. If the extrapolation is unreliable, then we should confine ourselves with the calibration area of our model. Taking into account that the highest rain gauge altitude is 540m, we recommend the study region should not extend over 700m (fig. 6, bottom, the black line). Therefore the entire westernmost part of our region should be excluded the final map because of extrapolation uncertainty.

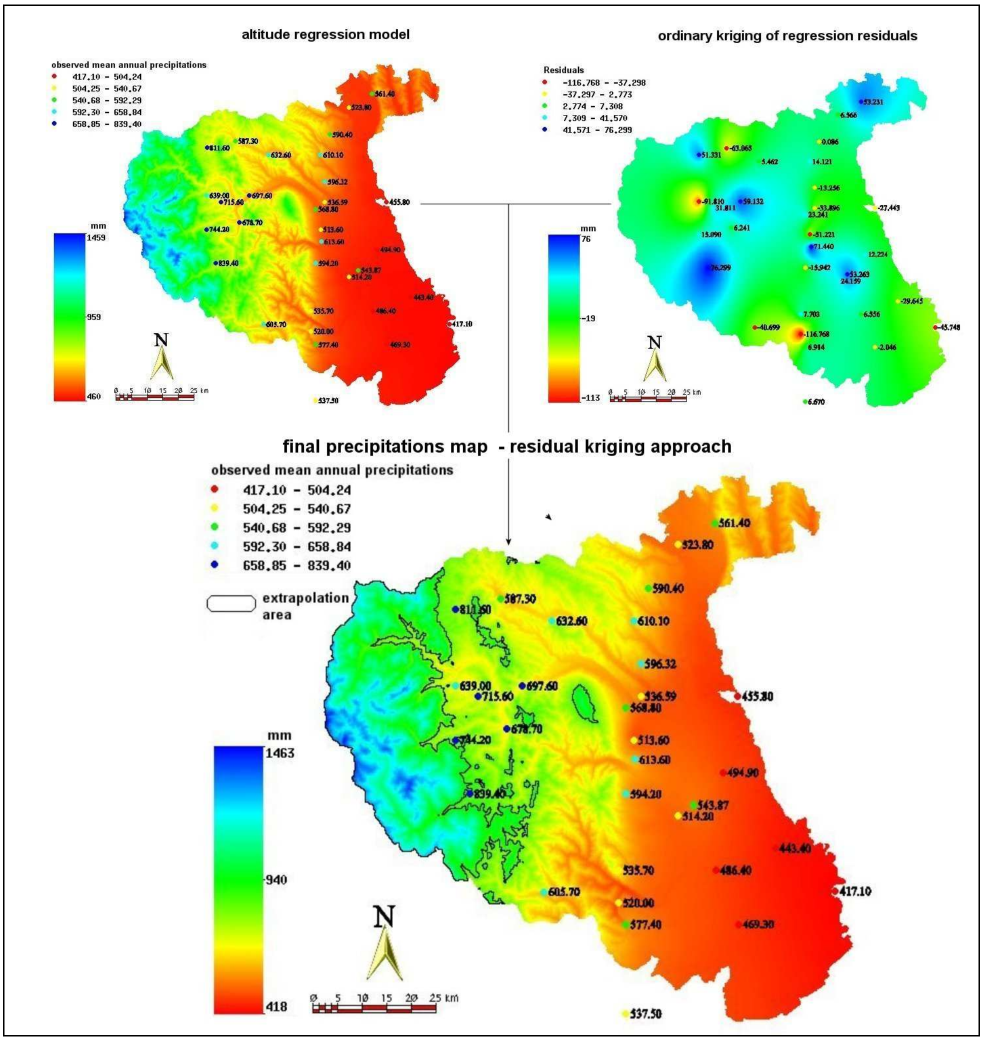
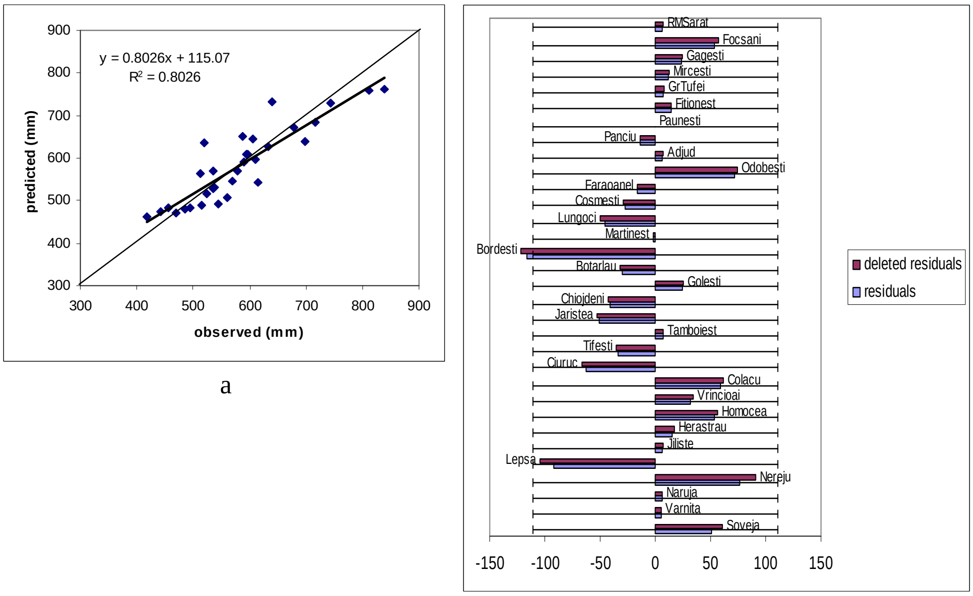


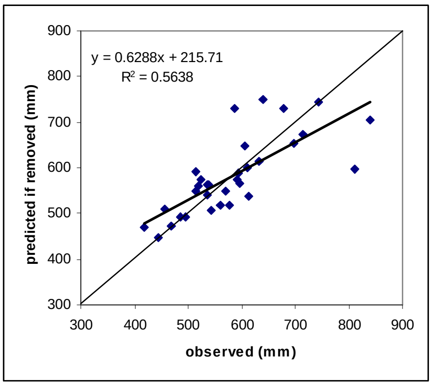
Fig.6 ***Residual kriging approach for mapping the mean annual precipitations***

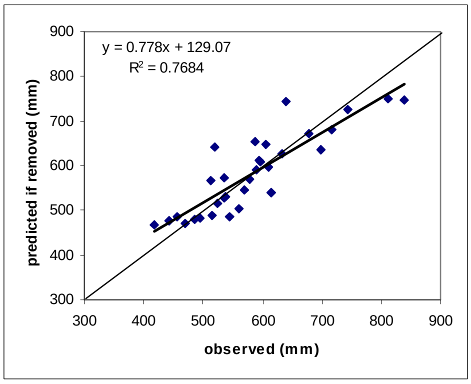
We come now to the final question of our analysis: which of the 2 models, ordinary kriging or residual kriging) is better suited for the mean annual precipitations spatialisation in our region? To answer it, we compared the cross-validation of the ordinary kriging output with the cross-validation of the altitude regression model and the standard error parameters of the ordinary kriging and residual kriging (fig. 8). We notice that there is a significantly better observed vs. predicted correlation for the altitude regression cross-validation and that all the standard error parameters (mean, minimum, maximum, standard deviation) have smaller values in the residual kriging approach, therefore indicating the superiority of this method.



a b

Fig. 7 ***Observed vs. predicted mean annual precipitations (a) and comparison between actual and deleted residuals (b) without the 2 outliers***

<table>
<tr><td>

cross-validation – ordinary kriging</td><td>

cross-validation – altitude regression</td></tr>
<tr><td>Standard error parameters of ordinary kriging of observed data
- mean: 12.663
- min: 0.017
- max: 50.634
- standard deviation: 13.568
a</td><td>Standard error parameters of ordinary kriging of regression residuals
- mean: 8.076
- min: 0.004
- max: 49.176
- standard deviation: 10.851
b</td></tr>
</table>

Fig. 8 ***Comparison between the performances of ordinary kriging (a) and residual kriging (b) for mapping mean annual precipitations***

5. CONCLUSIONS

When applying statistical methods for deriving digital spatial models of climatic variables one must take great care in identifying and assessing the sources of uncertainty. There are many such sources of different nature which can easily mislead us towards wrong unrealistic conclusions. Our article deals with the georeference errors and data errors / uncertainty, showing how such errors can affect the quality of our spatial models. The ordinary kriging or cokriging approaches are often used for deriving precipitation fields. However, our analysis shows that combining the regression and ordinary kriging in a residual kriging approach leads to better results, at least for mean annual precipitation values.

REFERENCES

Dobesch, H., Dumolard, P., Dyras, I., (editors, 2007), *Spatial Interpolation for Climate Data. The Use of GIS in Climatology and Meterology*, ISTE, 320 pp.

Hengl, T., (2007), *A Practical Guide to Geostatistical Mapping of Environmental Variables*. JRC Scientific and Technical Research series, Office for Official Publications of the European Comunities, Luxembourg, EUR 22904 EN, 143 pp.

Patriche, C.V., (2005), *Aportul metodelor statistice de interpolare la ameliorarea spațializării parametrilor climatici*, Memoriile Secțiilor Științifice, seria IV, tom XXVIII, Edit. Academiei Române, p. 93-107.

Patriche, C.V. (2007), *About the influence of space scale on the spatialisation of meteo-climatic variables*, Geographia Technica, Nr. 1 / 2007, Cluj University Press.