GEOSTATISTIC MODELING OF TEMPERATURE DATA FOR AGRICULTURAL LAND EVALUATION

Iuliana Cornelia Niculiță, Ionuț Minea
University "Al.I.Cuza" Iaşi, iuliananiculita@geomorphologyonline.com

Abstract. For low spatial resolution climate data (in Romania case) and relatively small study areas (used in land evaluation) classical methods of interpolation of climate data does not give significant spatial resolution results. To illustrate the use of geostatistic in solving this issue, we have compiled a database of annual average temperatures for the period 1947-2010 in Plateau of Moldavia, further used for applying kriging combined with regression of residues. Using this method and additional data for modeling multiple regression, we have been able to obtain a statistically and spatial valid resolution for an area with a surface of 949.3 km² located in the proximity of Iasi.

Keywords: *regression-kriging, Moldavian table land, DEM, diffuse radiation*

1. Introduction

Agricultural land evaluation uses a set of terrain data as an input for estimation of capability/favorability for different land uses and crops (ISSA, 1987). Climate data is essential in this approach, because controls both capability/favorability estimation. A good estimation of climatic data is needed, for obtaining good estimation of land suitability.

Most used climatic data for input in agricultural land evaluation methods are mean annual temperature and precipitation quantity. Mean seasonal and other climate parameters can also be used. In our approach we focused only on temperature, but the methodology, results and conclusion can also apply for the other climate data (with proper adaptation).

Spatial modeling of climate data involves geostatistics at any scale and type of data. The most used approaches are splines (ANUSPLIN – Hutchinson, 1995), spline with tension (Hofierka et. al., 2002) and RK (regression-kriging). RK (Hengl et. al., 2004, 2007, Patriche, 2009) is considered to be the best estimator because comprise a regression and the adding of the kriged residuals, giving statistical significance and spatial detail.

We focused on the Romanian climate datasets, because their spatial and temporal resolution needs intensive modeling. Our area for study concerning agricultural land evaluation has 949.3 km² and cover 15 villages north to south of Bahlui valley, between Târgu Frumos and Iaşi (described in Tanasă et. al., 2010). We want to obtain for this area a grid estimation of temperature which to be spatialised as much as possible.

We apply a regression-kriging approach and because the area is quite small and the climate data is sparse, we set on obtaining most spatial detailed modeling. This is why we try to obtain a big number of observations of the dependent data (temperature) and to add continuous spatial independent data, as much as possible.

For our study area WORLDCLIM data (Hijmans et. al., 2005) is too coarse resolution to be used.

2. Data and methods

As climatic data we used the mean annual temperature of 30 meteorological stations from Romania and 3 stations from Moldavian Republic, covering the Moldavian table land and neighbor areas. The sources of data were CMRM, ANM and GSOD database.

The data for the used meteorological stations is presented in Table 1. In Figure 1 we show the temporal coverage of different sources of temperature data. To obtain mean annual temperatures for 1947 – 2010 period for all the stations, we used linear regression of the closest station which gave the best correlation (evaluated using adjusted R^2, F test and t test p-value) for the overlapping period with data. All simple linear regressions have significance at a 0.01% level. GSOD database has missing daily data, so this data was used only where CMRM data was not available. Comparing GSOD and CMRM averages for overlapping years, shows that GSOD data averages with missing days have differences under 0.05°C, compared with CMRM data.

Statistical analysis of temperature data was performed in R (Itaka and Gentleman, 1996). We used *base* packages for input/output, manipulation of data, descriptive statistics and linear modeling, *sp* (Pebesma and Bivand, 2005, Bivand et. al., 2008) and *gstat* (Pebesma, 2004) for regression-kriging. *Gstat* package read *.asc* files, which are used in the regression kriging modeling.

The independent data for the regression was altitude, latitude, longitude and terrain modeled diffuse radiation. Meteorological station altitude, latitude and longitude were obtained from GSOD metadata and from topographic 1:25 000 scale maps.

Diffuse radiation data was modeled using a terrain based radiation model implemented in SAGA GIS (Boehner and Antonic, 2009). We modeled direct, diffuse and total radiation (kWh/m²/year), as a mean for a 365 days year, with one hour steps for every day, with latitude and longitude position for every pixel, a lumped atmospheric model with 70% transmissivity and visible sky grid. For computation efficiency of radiation modeling we used a SRTM DEM (NASA, version 2) resampled to 250 m (with B-spline in SAGA GIS) in Stereo 70 projection (reprojected from Geographic WGS84 made with SAGA GIS).

The used digital elevation model for our area of study, used in regression-kriging was obtained from contours (25k topographic maps) and height points (5k topographic maps), by interpolation with Thin Plate Splines function from SAGA GIS

2.0.7 (www.saga-gis.org), with 11 levels for a smooth generalization and outliers removal.

Table 1. Data for the used meteorological stations

No.	Meteo. station	Latitude (decimal degrees)	Longitude (decimal degrees)	Alt. (m)	T. (°C)	Rad. (kWh)	Country
1.	Rădăuți	47.838096	25.892051	388	7.40	913.49	RO
2.	Dorohoi	47.967000	26.400000	241	8.70	922.40	RO
3.	Darabani	48.195121	26.575084	247.5	8.20	920.93	RO
4.	Suceava	47.633134	26.242143	366	7.80	914.89	RO
5.	Suceava (Salcea)	47.680725	26.360156	400	7.70	912.58	RO
6.	Botoşani	47.735879	26.647139	160	9.30	922.48	RO
7.	Avrămeni	48.013624	26.960534	241	8.40	922.07	RO
8.	Rădăşeni	47.467000	26.250000	350	8.40	913.55	RO
9.	Răuseni	47.564173	27.208766	61	9.10	932.89	RO
10.	Ştefăneşti (Stânca)	47.832469	27.221309	105	9.00	925.58	RO
11.	Târgu Neamț	47.212376	26.380765	385	8.50	913.04	RO
12.	Strunga	47.158478	26.948730	280	8.50	919.56	RO
13.	Cotnari	47.358548	26.927209	285	9.20	919.17	RO
14.	Podu Iloaiei	47.214041	27.256401	82.5	9.50	931.30	RO
15.	Iaşi	47.171189	27.629868	102	9.70	930.54	RO
16.	Piatra Neamț	46.933923	26.391083	366	8.80	913.97	RO
17.	Roman	46.969338	26.913394	217	8.80	923.34	RO
18.	Bacău	46.532148	26.914067	181	9.20	925.52	RO
19.	Negreşti (Vaslui)	46.838328	27.443695	132	9.10	927.68	RO
20.	Vaslui	46.646347	27.715986	117	9.50	928.00	RO
21.	Adjud	46.104974	27.171934	103	9.70	930.29	RO
22.	Plopana	46.685701	27.208545	262.5	9.00	914.96	RO
23.	Onceşti	46.475645	27.260819	203	9.90	921.47	RO
24.	Bârlad	46.233288	27.645977	170	9.80	926.09	RO
25.	Huşi	46.677813	28.102103	85	10.30	931.21	RO
26.	Focşani	45.687777	27.201305	51	10.60	933.73	RO
27.	Tecuci	45.833333	27.400000	60	10.10	933.08	RO
28.	Balinteşti	46.023769	27.923046	120	9.70	925.99	RO
29.	Galați	45.473163	28.033801	65	10.80	932.50	RO
30.	Măicăneşti	45.501849	27.493578	15	10.80	935.73	RO
31.	Bălți	47.783000	27.950000	102	8.90	929.89	MD
32.	Lozova	47.333000	28.083000	232	8.90	915.71	MD
33.	Chişinău	46.917000	28.933000	122	9.80	931.68	MD

This approach and a pixel size of 12.5 m were used after Hengl, 2006 guidelines for eliminating as good as possible the interpolation errors. The DEM was further hydrological preprocessed, by filling sinks and by imposition of the hydrographic network extracted from 1: 25 000 scale topographic maps.

Latitude and longitude grid data for regression kriging were obtained with Natural Neighbour interpolation algorithms (in SAGA GIS) from .shp file containing extent latitude and longitude data, as attributes.

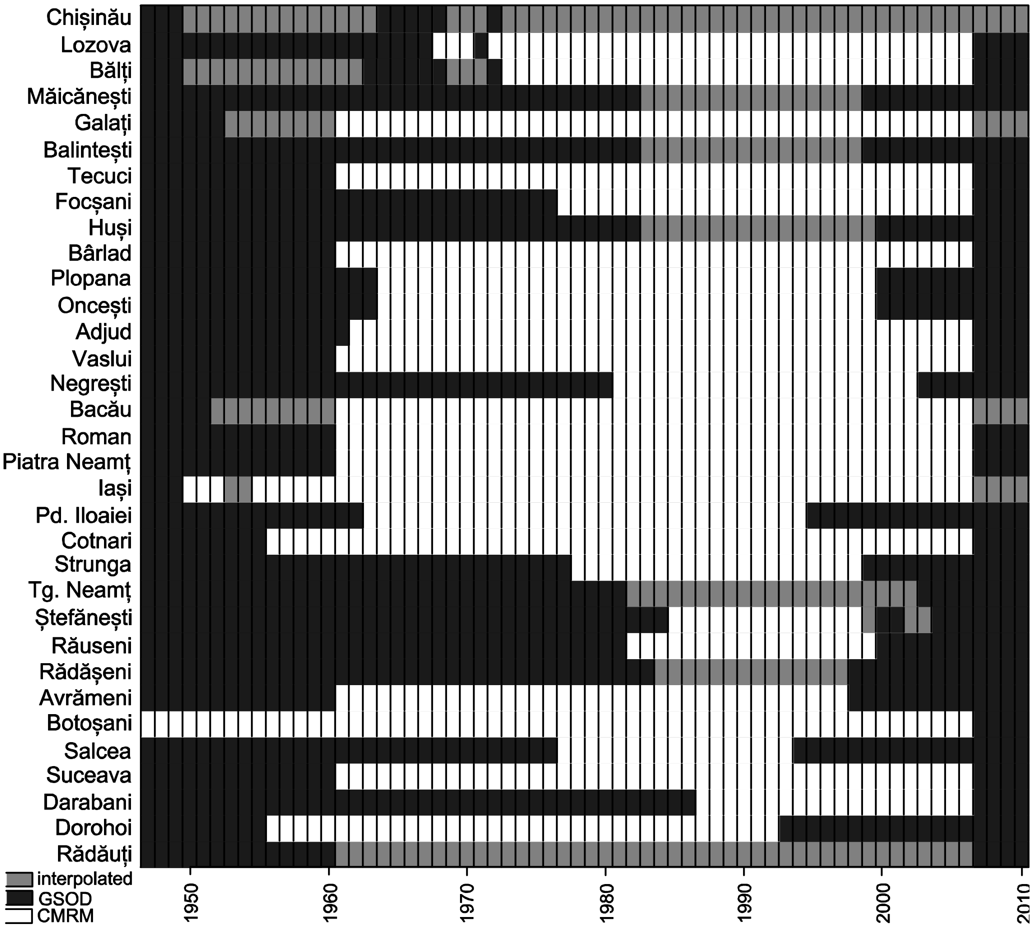

Figure 1 Temporal coverage of meteorological data from different sources

Diffuse radiation grid data for our study area was modeled also in SAGA GIS, with the mentioned implementation, but for computational efficiency we used 4 hour day steps (24/0, 4, 8, 12, 16, 20).

3. Results and discussions

The number of temperature database observations (33) is under the limit expressed by Hengl et. al., 2007 (50) for application of RK. Compared with other papers describing moldavian temperature datasets (Patriche, 2009) our number of observation is the biggest.

Although by using multiple regression, the significance of the modeling decrease slowly (regression only with altitude and latitude gives adjusted R^2 of 0.8716, while altitude, latitude, longitude and diffuse radiation gives adjusted R^2 of 0.8695), the regression-kriging approach introduce a bigger spatial detail for our small area of study. The collinearity of the independent variables, was tested using a principal component analysis, showing that the variables are not dependent.

The obtained regression equation is:

temp. *= 38.250857 - 0.003916 ** ***alt.*** *- 0.563050 ** ***lat.*** *+ 0.138249 ** ***long.*** *- 0.006079 ** ***diffuse rad.*** (1)

The adjusted R correlation coefficient is 0.8695, the partial regression coefficients being 0.87 for altitude, 0.65 for latitude, 0.46 for longitude, and 0.63 for diffuse radiation. For the F test, the F statistic is 54.3, for (4, 28) freedom degrees, while for the 0.01 level we need a value > 14.020, showing that the data is conformal with the nested multiple regression models. The *p* value of the Student test is 8.58e-13, rejecting the null hypothesis, so the test is statistically significant.

Descriptive statistic of residuals is shown in Table 2. Compared with the regression modeling of Patriche, 2009, which used 12 observations, we obtain bigger residuals, but we argument that we need a final detailed spatial modeling, rather than a perfect statistic model.

Table 2. Descriptive statistics of regression residuals

No.	Meteo. station	Residual	No.	Meteo. station	Residual	No.	Meteo. station	Residual
1.	Rădăuți	-0.4232	12.	Strunga	-0.2379	23.	Onceşti	0.4446
2.	Dorohoi	0.3577	13.	Cotnari	0.5949	24.	Bârlad	0.0537
3.	Darabani	-0.0215	14.	Podu Iloaiei	0.0489	25.	Huşi	0.4392
4.	Suceava	-0.2646	15.	Iaşi	0.2449	26.	Focşani	0.1886
5.	Salcea	-0.2351	16.	Piatra Neamț	0.3155	27.	Tecuci	-0.2257
6.	Botoşani	0.4768	17.	Roman	-0.2632	28.	Balinteşti	-0.3989
7.	Avrămeni	0.0045	18.	Bacău	-0.2372	29.	Galați	0.2000
8.	Rădăşeni	0.1699	19.	Negreşti	-0.4168	30.	Măicăneşti	0.1147
9.	Răuseni	-0.2219	20.	Vaslui	-0.2193	31.	Bălți	-0.2589
10.	Ştefăneşti	-0.0447	21.	Adjud	-0.2898	32.	Lozova	-0.1078
11.	Tg. Neamț	0.2424	22.	Plopana	-0.1365	33.	Chişinău	0.1068
Min.	-0.4232	Mean	0.00	Max.	0.5949		Std. dev.	0.2852

The introduction in regression of terrain modeled radiation, extend the empirical approach of Xin and Chenchao, 2008 and Patriche, 2009, which use the insolation and aspect, for correction of modeled temperature data.

Regression-kriging was performed in two steps. First we applied the regression equation over the independent variables grids. Then we interpolated by ordinary kriging (OK), using a global search and the parameters from Table 3, temperature residuals. Semivariogram plot of temperature residuals is represented in Fig. 3.

Table 3. Parameters of the variogram modeling

Variogram model	Nugget	Sill	Range
exponential	0	0.09	10 451.5

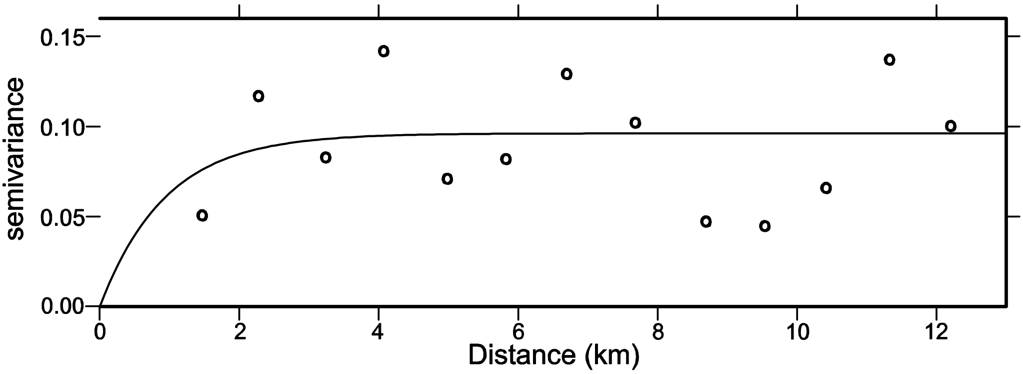

Figure 3 Semivariogram plot of residuals

The spatial results of the temperature modeling are represented in Fig. 4. The descriptive statistic for the layers used in regression kriging and for the final results are represented in Table 4.

Table 4. Descriptive statistics of regression kriging results

Variogram model	Multiple regression	OK of resids	RK of temp.	Altitude (m)
Min.	8.048	-0.0827	8.097	42.76
Mean	8.955	-0.0347	8.921	121.1
Max.	10.47	0.0500	10.40	347.3

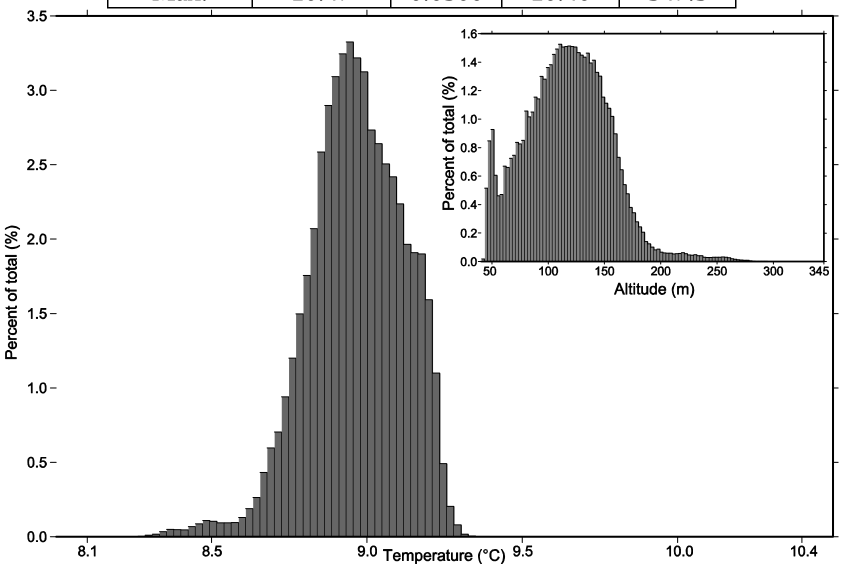

Figure 4 Histogram of altitude and RK temperature modeling

Conclusions

Geostatistic provide the framework for climate data modeling to obtain a spatial estimation of data available in a climate monitoring network. Especially in areas with sparse spatial coverage, and for relative small areas, the simple spatial interpolation methods (like spline), don't give enough spatial resolution. Geostatistic methods like regression-kriging, coupled with additional topographic data, can be used to increase the spatial resolution. We present a usable database of temperature data for Moldavia table land and the application of regression-kriging to model temperature for a small area near Iaşi city.

Acknowledgments

This work was supported by the European Social Fund in Romania, under the responsibility of the managing Authority for the Sectoral Operational Programme for Human Resources Development 2007-2013 [grant POSDRU/88/1.5/S/47646 and grant POSDRU/89/1.5/S/63663 „Rețea transnațională de management integrat al cercetării postdoctorale în domeniul Comunicarea științei. Construcție instituțională (şcoala postdoctorală) şi program de burse (CommScie)"].

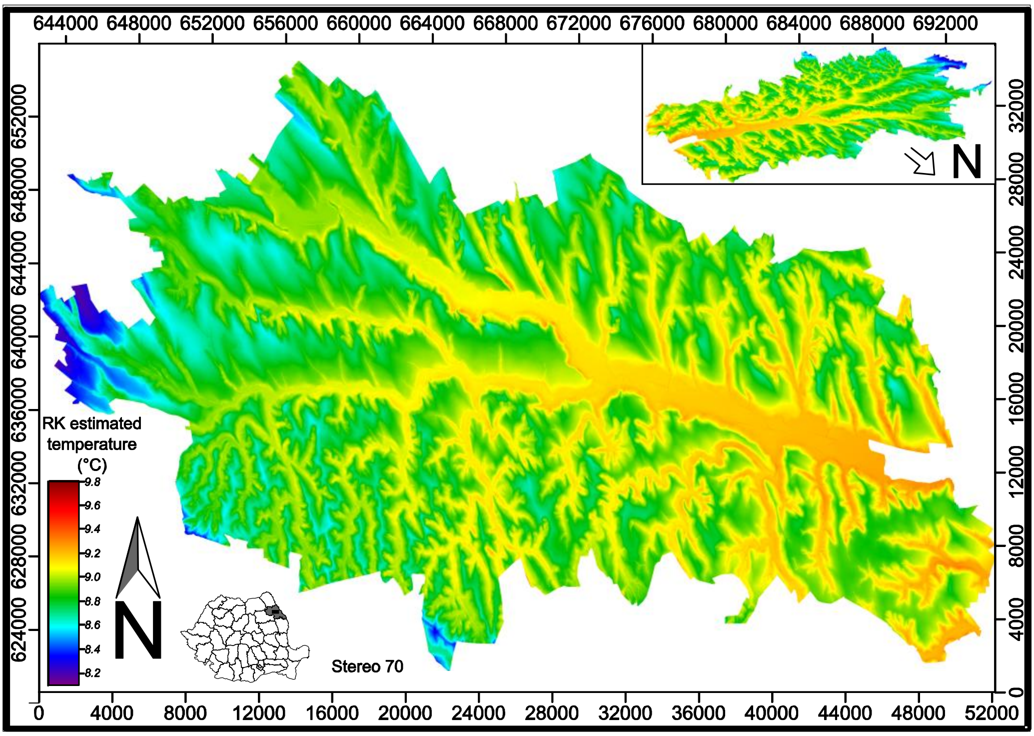

Figure 5 Spatial results of the RK of temperature for our study area

References

Bivand R.S., Pebesma E.J., Gomez-Rubio V., 2008. *Applied spatial data analysis with R*, Springer.

Boehner J., Antonic O., 2009. *Land surface parameters specific to topo-climatology*, in Hengl T. and Reuter H.I., eds., *Geomorphometry: concepts, software, applications*, Elsevier.

CMRM meteorological tables, (Moldavian Regional Meteorological Centre).

Coordinate calculator - http://transition.fcc.gov/mb/audio/bickel/DDDMMSS-decimal.html. (accessed 01.06.2011)

Hengl T., Heuvelink G.B.M., Stein A., 2004. *A generic framework for spatial prediction of soil variables based on regression-kriging*, Geoderma, 120:75-93.

Hengl T., 2006. *Finding the right pixel size*, Computers & Geosciences, 32:1283-1298.

Hengl T., Heuvelink G.B.M., Rossiter DG., 2007. *About regression-kriging: from equations to case studies*, Computers & Geosciences, 33:1301-1315.

Hijmans R.J., Cameron S.E., Parra J.L., Jones P.G., Jarvis A., 2005. *Very high resolution interpolated climate surfaces for global land areas*, International Journal of Climatology, 25:1965-1978.

Hutchinson M.F., 1995. *Interpolating mean rainfall using thin plate smoothing splines*, International Journal of Geographical Information Systems, 9(4):385-403.

Ihaka R., Gentleman R.R., 1996. *A language for data analysis and graphics*, Journal of Computational and Graphical Statistics, 5:299-314.

Institute for Soil Science and Agrochemistry, 1987 – *Methodology of pedologic studies elaboration, IInd part – Elaboration of pedological studies for various scopes*, 349 p., Bucharest. (in romanian)

Hofierka J., Parajka J., Mitasova H., Mitas L., 2002. *Multivariate interpolation of precipitation using regularized spline with tension*, Transaction in GIS, 6(2):135-150.

National Climatic Data Center, NOAA, 2006. *Global Summary of the Day (GSOD)*, http://www.ncdc.noaa.gov/cgi-bin/res40.pl?page=gsod.html. (accessed 01.06.2011)

Patriche C.V., 2009. *Statistical methods applied in climatology*, Terra Nostra Press, Iaș i. (in romanian)

Pebesma E.J., 2004. *Multivariable geostatistics in S: the gstat package*, Computers & Geosciences, 30:683–691.

Pebesma E.J., Bivand R.S., 2005. *Classes and methods for spatial data in R*, R News 5(2), http://cran.r-project.org/doc/Rnews/.

Tanasă I.C., Niculiță M., Roşca B., Pîrnău R., 2010. *Pedometric techniques in spatialisation of soil properties for agricultural land evaluation*, Bull. UASVM Agric. 67:274-278.

SRTM v. 2, NASA, http://dds.cr.usgs.gov/srtm/.

Xin, Y., Chenchao X., 2008. *Terrain-based revision of an air temperature model in mountain areas*, in Zhou Q., Lees B., Tang G., eds., Advances in digital terrain analysis, Springer.

ANM meteorological data provided through research grant CEEX 775/2006.