ABOUT THE INFLUENCE OF SPACE SCALE ON THE SPATIALISATION OF METEO-CLIMATIC VARIABLES

C. V. Patriche[1]

ABSTRACT
The article approaches some problems related to the in influence of the space scale on the quality of statistically based spatial models of meteo-climatic variables. The neighborhood information issue is discussed, meaning the extent to which the information from the area surrounding a point of measurement is more relevant for spatialisation than the strictly local information. Then we approach the problem of spatialisation in heterogeneous regions and the applicability of the models at different scales. Finally, we tackle the problem of the outliers, the presence of which could lead us to incorrect interpretations.

The elaboration of statistically based spatial models for meteo-climatic variables requires insight on some aspects, which could help us improve the models or avoid errors. We discuss in this article some of these aspects, related to the influence of the space scale on the quality of the spatialisations.

1. THE NEIGHBORHOOD INFORMATION

Sometimes, the information from the area surrounding a meteorological station / rain gauge is more relevant for the parameter being analyzed than the strictly local information associated to the stations' locations.

The simplest way to account for the neighborhood information is to calculate the mean predictor values for the surrounding area and to test which size of this area is better correlated to the analyzed parameter. In GIS, this is done by filtering the predictors using low-pass filters. Then the predictors mean values can be automatically extracted from these filtered grids.

Previous research has proved the usefulness of this technique mainly for temperature and radiation related variables (*Lhotellier R*, 2005, *Patriche C. V., Lhotellier R*, 2006, *Patriche C. V.*, 2006, *Paul and David 2006*). From figure 1 we may notice that the maximum (best) correlation between different climatic parameters and altitude is, in most cases, associated with a filtered DEM – see (Imecs 2006) - and not with the raw (local) altitudes. These studies have shown that the increase of the explained variance by using low-pass filtered predictors instead of the raw predictors may be as high as 18%, which justifies the use of this technique for the spatialisation of climatic variables.

The optimum moving window size, which is the size associated with the best correlation coefficient, varies greatly from one parameter to another and from one time frame to another (figure 2).

[1] *Romanian Academy, Department of Iasi, Geography Group*

Example from the French Alps for mean monthly minimum temperature (*Patriche C. V., Lhotellier R.*, 2006)

Example from the Moldavian Plateau (*Patriche C. V.*, 2006)

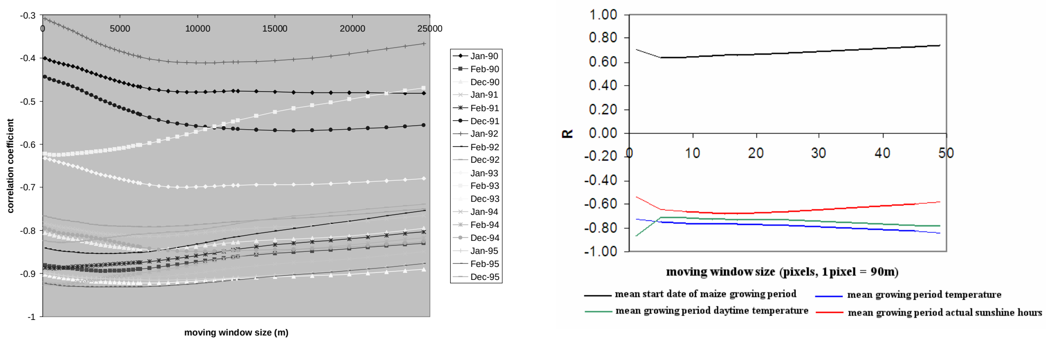

Fig. 1. ***Variations of altitude correlations with different climatic parameters according to the moving windows size***

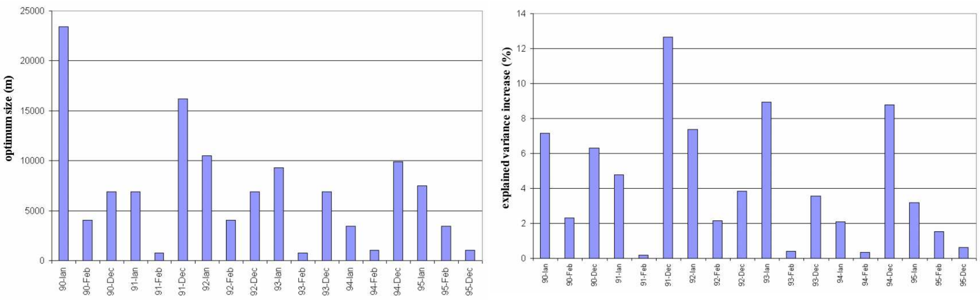

Fig. 2. ***The optimum moving window size (left) and the increase in the explained variance (right) caused by using the filtered DEMs instead of the raw altitudes. Example from the French Alps for mean monthly minimum temperature (Patriche C. V., Lhotellier R., 2006)***

The example shown in figure 3 reveals that mean maize growing period temperature in the Moldavian Plateau is correlated the best with the mean altitude values from a surrounding area of 4410 x 4410m, the increase of the explained variance, with respect to the local altitude correlation, being of 13%.

Apart from using the mean predictor values, other moving windows operations may also prove useful for climatic parameters spatialisation, such as the range or standard deviation (e.g. elevation standard deviation). Much more complicated techniques the principal components analysis in order to account for the neighborhood information, such as the Aurelhy method (*Benichou P., Le Breton O.*, 1987) designed for precipitations spatialisation in complex terrain.

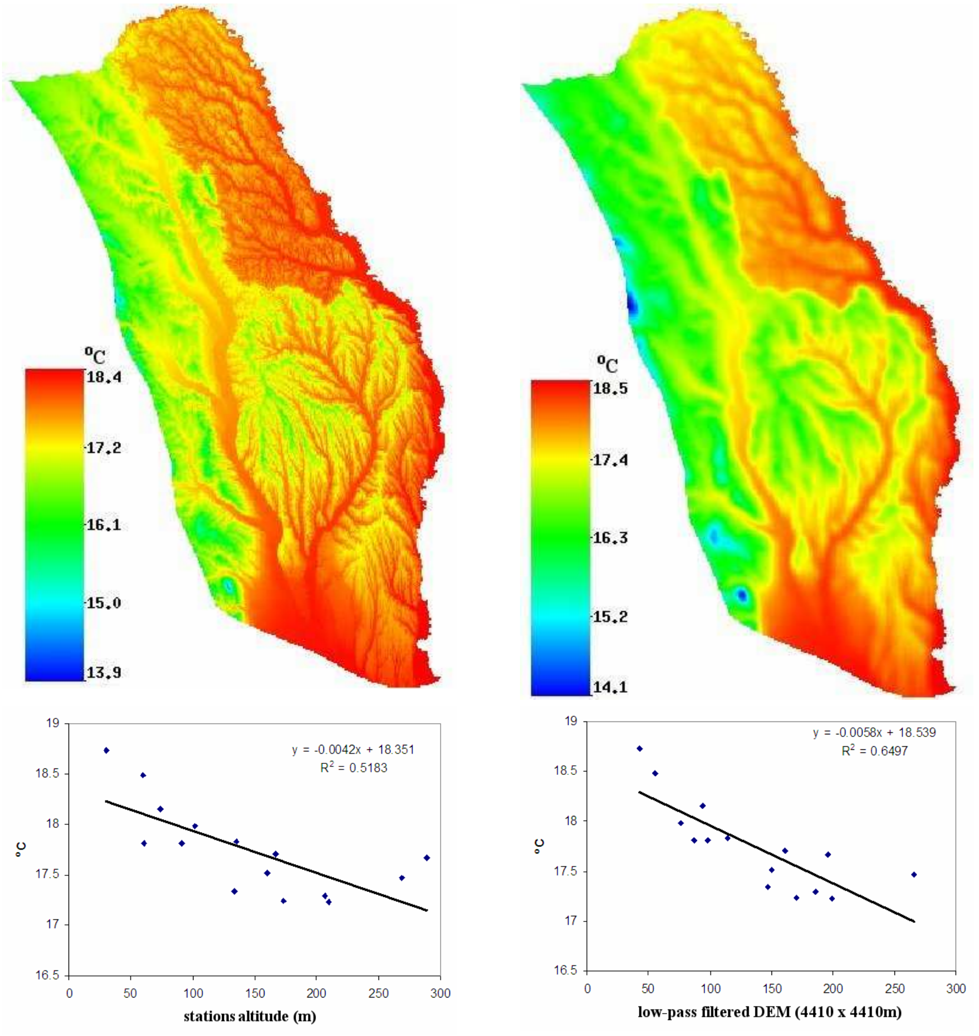

Fig. 3. ***Comparison between the spatialisations of the altitude – temperature relationship using the stations altitudes (left) the filtered 4410 x 4410m DEM (right). Example from the Moldavian Plateau for the mean maize growing period temperature***

2. THE PROBLEM OF HETEROGENEOUS REGIONS

Another issue related to the influence of space scale on climatic space models is that of the heterogeneous regions. Generally, as the scale decreases, the area of investigation becomes larger and therefore more heterogeneous. At local scale, the elaboration of statistically based spatial models is hampered by terrain homogeneity, besides the sparse station network. At first, the decrease of scale is useful for spatialisation because the terrain begins to reveal more and more of its characteristics, becoming therefore able to explain better the spatial distribution of climatic parameters. The problems occur when the decrease of the space scale determines the inclusion within the interest area of a region where the predictor – predictand relationship changes significantly. For such heterogeneous regions, one cannot apply a single regression equation to models the climatic fields. These regions must be first divided into smaller, less heterogeneous, areas, for which the predictor – predictand relationship remains at the same parameters.

An example is shown in figure 4 for the relationship between the mean annual temperature and 3 predictors: altitude, latitude and longitude. At continental scale, the territory of Europe is very heterogeneous. We may notice that the altitude – temperature relationship changes form one region to another to such an extent that a single regression equation for the whole European territory cannot be constructed. A region like the Alps displays a very good altitude – temperature correlation, while the temperature variation within the flat relief of the Russian Plain is statistically independent of the altitude, as temperature inversions are frequent. Here, the latitude comes forward to explain a good part of the temperature spatial distribution.

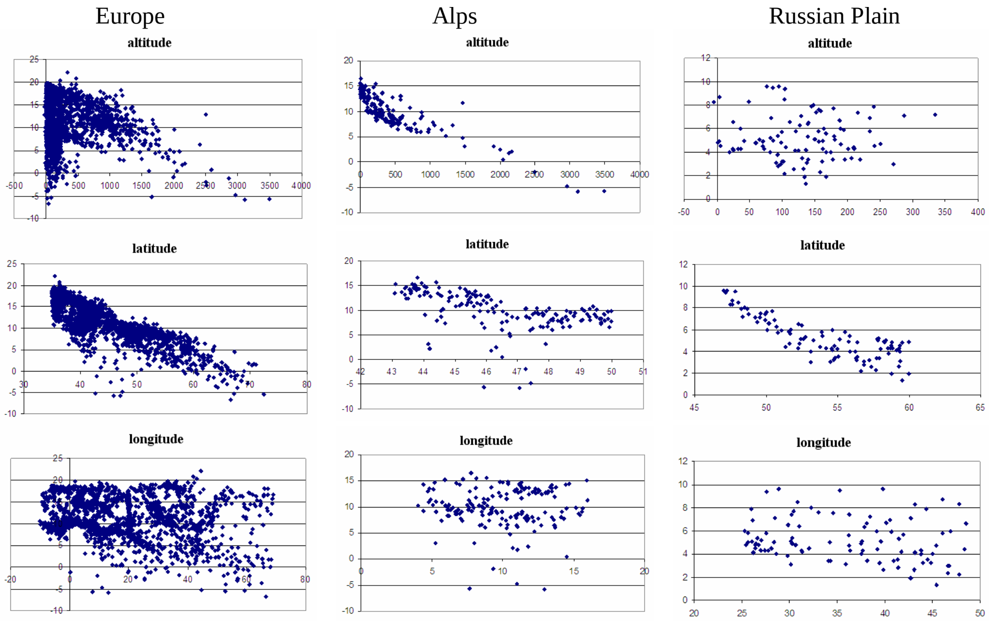

Fig. 4. ***Changes of the relationships between the mean annual temperature and the altitude, latitude and longitude for different regions in Europe***

If we deal with a heterogeneous region, then we face the problem of dividing it into sub-regions, which must not be too homogeneous because then we won't the capable of grasping the real predictor – predictand relationship. One way to do that is to examine the change in the regression parameters and stations' residues as we expand (or contract) our study region. Then we can establish the limits of our sub-region, which corresponds to the most stable regression model (maximum correlation, minimum residues). Another way to deal with heterogeneous regions is to apply regression as a local interpolator, but this is frequently hampered by the scarcity of the stations network.

3. THE MODELS' APPLICABILITY AT DIFFERENT SPACE SCALES

This issue is tightly related to the previous one. The question we ask here is to what extent a spatial model elaborated for a certain scale is applicable at a different scale. Figure 5 shows an example along this line, where a mean annual stepwise regression model with altitude and latitude as predictors is applied for a much smaller region, respectively the Moldavian Plateau from eastern Romania. We may notice the different temperature gradients revealed by the two regression equations. While the vertical gradient values are very similar, the latitudinal gradients differ considerably. But due to the small latitudinal extent of the Moldavian Plateau (about 2^{o}), the errors induced by the different latitudinal gradients are small. Nevertheless, the European model tends to overestimate higher temperature values, the maximum estimated mean annual temperature being 12.8^{o}C, which is not found on the Romanian territory. The mean difference between the two spatial models is –1.6°C, which is significant for this parameter. Consequently we may draw the conclusion that, in our example, the continental scale temperature model is not applicable to the regional scale of the Moldavian Plateau.

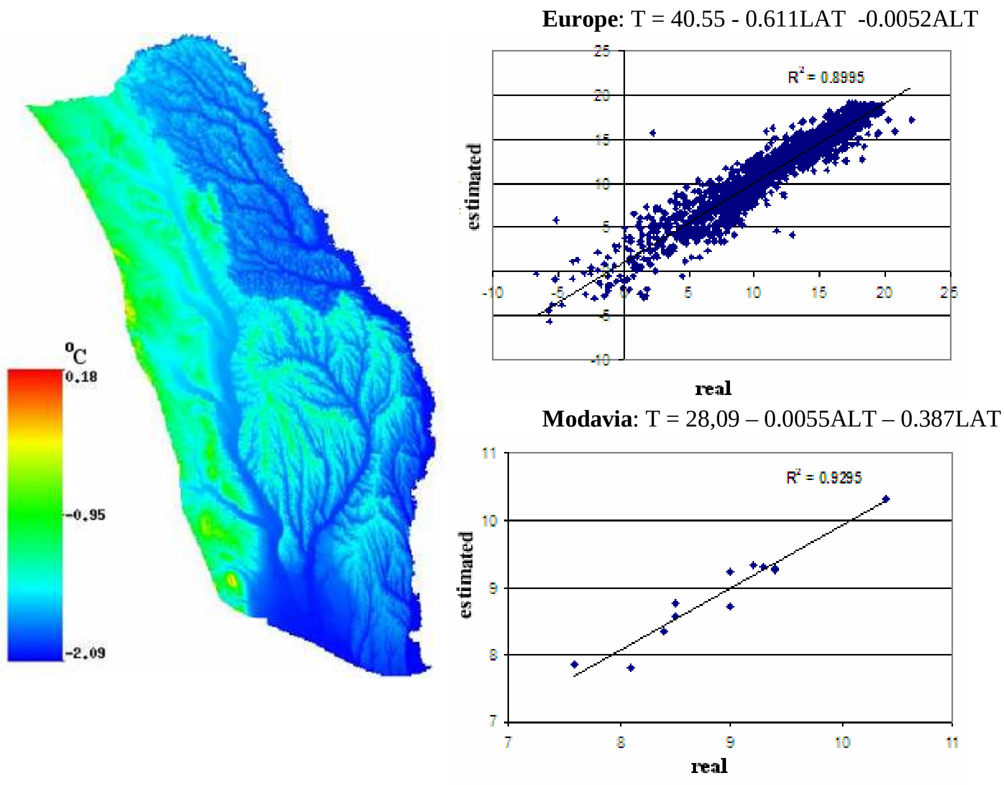

Fig. 5. ***The difference between the mean annual temperature spatialisations as functions of altitude and latitude, using the Moldavian Plateau stations and all Europe stations.***

As previously stated, the optimum interpolation region could be identified by noticing the changes of the regression parameters as we expand or contract our region. Figure 6 shows what happens when we expand our plateau region into a mountainous area. The additions to our stations sample of 2 outside higher altitude stations changes significantly the mean annual temperature – altitude regression model. The vertical gradient drops from 0.57°C to 0.47°C and the relation seems to improve itself by the significant increase of the explained variance (from 36.5% to 95.1%). However, the inspection of the residues for the stations situated within the plateau area shows that the temperature values are better estimated using the lower explanative regression model derived form the plateau stations sample only. The apparent improvement of the regression by inserting the 2 outside stations, situated in different climatic conditions, with much lower temperature values, is caused by the effect of “attraction” of the regression line by these points with very different values.

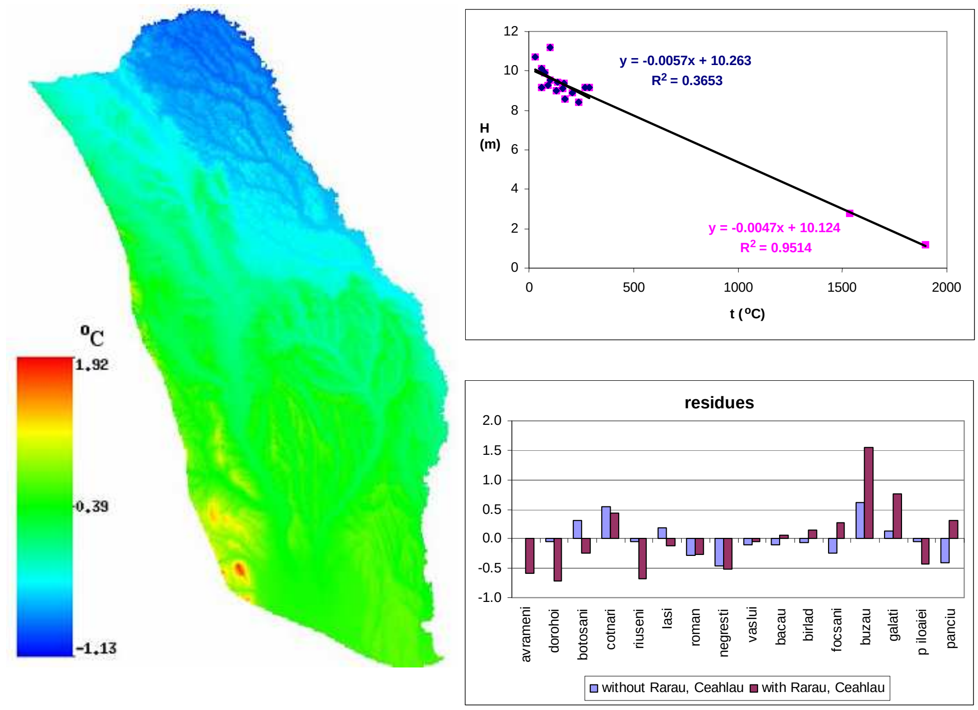

Fig. 6. ***The difference between the mean annual temperature spatialisations as functions of altitude with and without 2 outside higher altitude stations***

Therefore we may conclude that the expansion of our plateau region into a mountainous one creates a heterogeneous region for which a single altitude regression model cannot be applied. If the plateau area is our region of interest, we should confine ourselves with the lower predictive regression model derived from inside stations only.

1. THE OUTLIERS PROBLEM

The regression models can be negatively influenced by the presence of values evading the spatial variation rules stated by the models (outliers). An example is shown in figure 7 for the mean annual temperature – altitude regression from eastern Romania. The charts display the variation of the correlation coefficients as new stations is progressively included, starting from a sample of 5 stations situated within the Moldavian Plain. We may notice that the inclusion of Cotnari station decreases significantly the correlation, keeping it at lower levels even after the insertions of several other stations. This is due to the fact that Cotnari station is situated in a föehnization area and therefore the temperatures are higher and the precipitations are lower than one would expect for its altitude. If we eliminate this station the regression models improve significantly.

with Cotnari station | without Cotnari station

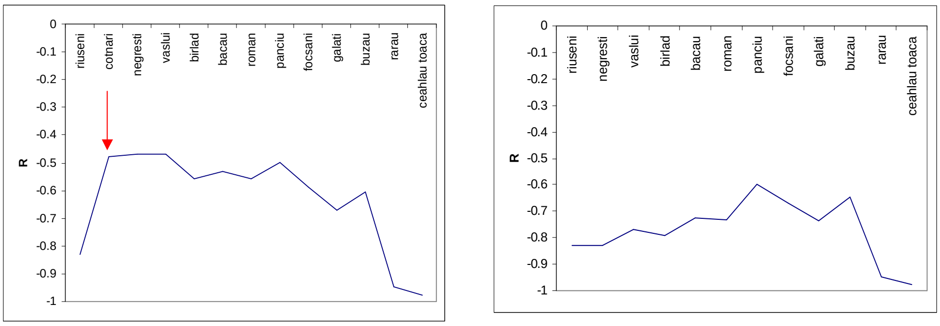

Fig. 7. ***Variation of mean annual temperature – altitude correlation coefficients as new stations are progressively included***

with Cotnari station | without Cotnari station

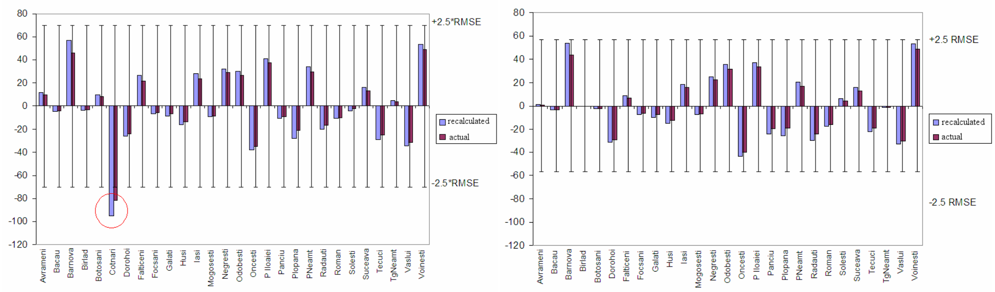

Fig. 8. ***Identification of outliers by comparing the residues derived from the regression models with and without Cotnari station (cross-validation). Example for mean annual precipitations within maize growing period in Moldavian Plateau***

But is it correct to eliminate this station? For the regression models, yes, unless we insert another predictor that would be able to account for this spatial anomaly. For the final temperature map, no, because the föehnization area would not be represented. The solution may be a residual kriging spatialisation, which adds the regression residues to the spatial trend. Another approach would be to use the regression as a local interpolator, as mentioned before. This, however, would require a dense stations network, which is not the case for our study region.

How can we identify an outlier? How great should a residue value be in order to regard the corresponding point as an outlier? An example is shown in figure 8 for the same Cotnari station, but for a different climatic parameter (mean annual precipitations). First, we should inspect the magnitude of the residues. If some value goes out the interval limited by ±2.5 RMSE[1] (equivalent in our case with the standard deviation of the residues), then it is possible that this value is an outlier. To establish that, one could compare the residues with those obtained by eliminating the suspect point (cross-validation). If the two regression models are stable, then the magnitude of the residues should be very similar.

[1] Root Mean Square Error

In our example, we notice that the difference between the actual (with Cotnari station) and the recalculated (without Cotnari station) is the greatest in the case of Cotnari station, which means that its exclusion from the model changes significantly the altitude – precipitation relationship.

REFERENCES

1. Benichou, P., Le Breton, O. (1987), *Prise en compte de la topographie pour la cartographie de champs pluviométriques statistiques*, La Météorologie, 1987, 7, 19, 23-35.
2. Imecs Z., (2006), *Precizia modelului SRTM*, Geographia Technica Nr.1/2006, pp.97-102, ISSN 1842-5135
3. Lhotellier, R. (2005) – *Spatialisation des températures en zone de montagne alpine*, thèse de doctorat, SEIGAD, IGA, Univ. J. Fourier, Grenoble, France, 350 p.
4. Patriche, C. V., Lhotellier, R. (2006), Possibilities of improving the regression models used for climatic parameters spatialisation. Study case: winter mean minimum temperatures in the French Alps, An. St. Univ. “Al. I. Cuza” Iaşi, s. II c., Geography, LII, pp. 28-39.
5. Patriche, C. V. (2006), *Modélisation de quelques variables agro-climatiques*, Actes du Colloque International: Observation et analyse des territoires ruraux, Simona Niculescu (Editor), Edit. Sedcom Libris, Iasi, pp. 102-120.
6. Paul, P., David, B.S., 2006**,** *Analysis of the historical precipitation sums of Sulina Station by means of power spectra in relation to Sibiu Station and NAO and SOI Indexes*, Geographia Technica Nr.2/2006, pp.99-104, ISSN 1842-5135