Using multivariate techniques for identification the atmospheric circulation patterns associated to heavy precipitation in Romania

STSM Scientific Report

Guest Scientist
Ioana Colfescu
National Meteorological Administration

Host Scientist
Professor Javier Martin-Vide
University of Barcelona,
Group of Climatology

15-22 February 2008
University of Barcelona, Group of Climatology

1. PURPOSE OF THE VISIT

The classification of atmospheric circulation is an issue of increasing interest among the climatologists as it allows a better understanding of the large scale mechanisms that generate extreme weather events such as heavy precipitation, heavy snowfalls, heat waves, etc. Most of these techniques are based on the use of multivariate statistics, principal component analysis (PCA) and clustering methods (CL). A large number of applications of these techniques including classification of weather types, climate regionalization, atmospheric circulation reconstructions, circulation patterns associated to climate extremes have been published (Esteban et al. 2005 and 2006; Romero et al. 1999).

During the STSM at the University of Barcelona we shared the experience of Spanish colleagues to identify the atmospheric circulation patterns associated to heavy precipitation in Romania – we used the methodology proposed by Esteban et al. (2005 and 2007).

The Principal Component Analysis (PCA) used for the SLP(sea level pressure) and g700 (geopotential height at 700 hPa) data to identify the principal modes of variability of these fields. PCA was based on S-mode data matrix, the variables being the grid points and the days being the observations. Then, the non-hierarhical K-means method was used to cluster the data according to the component scores obtained from PCA.

DESCRIPTION OF THE WORK CARRIED OUT DURING THE VISIT

2.Description of the data and methodology

2.1.Data

Daily mean data from 30 Romanian stations (Fig.1) were used to make the selection of the heavy precipitation days. The stations cover the intracharpatian area including also the mountain stations bordering the area
A heavy precipitation day was defined as that day when daily precipitation total exceeded the threshold of 15 mm. The precipitation threshold was established by calculating the 95% percentile for each station and then selecting the average value of that percentile over the region.

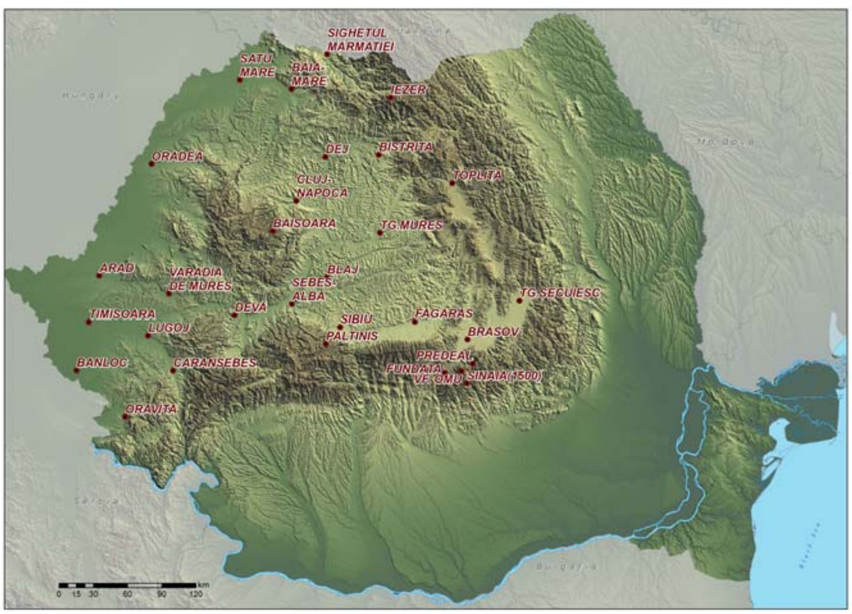


Figure1. The 30 Romanian (intracarphatian) stations used for the study

The period of our study covered 45 winter seasons (December –January-February) from 1961 to 2006. As large scale data we used daily means of SLP and geopotential heights at 700 hPa from the NCEP – NCAR reanalysis data . The data used encompasses the region of the Atlantic-European domain between 35–70^o N and 30º W–50º E, with a 2.5º spatial resolution .
The analysis we made consisted of two main steps: first was applied the PCA on the large scale data corresponding to that days when daily precipitation

exceeded the threshold and then was applied the k-means method of cluster analysis in order to obtain the circulation patterns associated to heavy precipitation .

2.2. METHODOLOGY - PCA

We first standardized the data, from the spatial point of view. This procedure deseasonalizes the observations, keeping the fields' intensities in the analysis, and maintains the daily temporal scale of the original data (Esteban et. al, 2005). Then we applied the PCA on a S-mode standardized matrix – the columns representing the grid points and the rows representing the days. The PCA was applied on the correlation matrix. The decision to choose the correlation matrix or the covariance matrix for PCA depends on how we wish the variance at each grid points are weighted. With the correlation matrix all the elements receive the same weight and only the structure, not the amplitude will influence the principal components so for this reason we choose to apply PCA on the correlation matrix .Then , the number of components explaining a significant portion of the total variance (Esteban et. al, 2005) was determined by means of the Scree test . Then the retained components were rotated using a varimax procedure.

2.3. METHODOLOGY - Clustering

Principal component analysis (PCA) is a widely used statistical technique for dimension reduction. *K*-means clustering is a commonly used data clustering for performing grouping of the data. In the study here the K-means method is used to cluster the observation but to decide the number of groups and centroids needed for the classification we use the spatial variation patterns established by the PC. We make this based on the fact that that principal components are the continuous solutions to the discrete cluster membership indicators for *K*-means clustering.

After applying the PCA we established to take a number of PCs of whose explained variance is more than 80%. Further, we consider these PCs in order to create the centroids of this groups and to establish the number of clusters. Considering this, observations with high score values for a certain component were selected (values higher than +1 for the positive phase, or lower than −1 for the negative phase, but being between +1 and −1 for the remainder of the PCs retained) . Using this technique, some components in their positive or negative phase were eliminated. The centroids were obtained from the average of the cases included into every group .Finally, the *K* - means produces the final classification of all the observations (days) with similar distribution of the SLP fields. We forced non-iteration in the *K* - means clustering method, as the centroids were well established by the first PCA grouping (reflecting a circulation pattern) because we are interested not so much in the convergence of the clustering algorithm (and not at all in changing the centroids) but in the spatial structure at the first step of the k-means analysis. In this way we obtained the clusters of the circulation patterns and eventually, we got synopticmaps of the atmospheric circulation groups for SLP /g700 with the same spatial distribution .

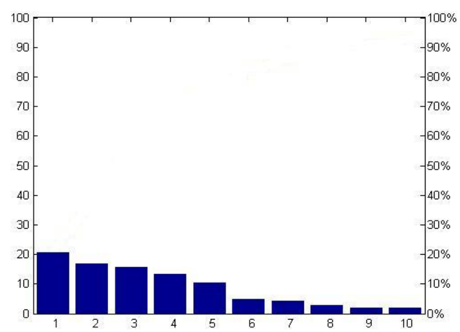
Eventually, based on the analysis of the results of the k-means method we did a regrouping of clusters into a smaller number of clusters to obtain the most representative circulation patterns.

DESCRIPTION OF THE MAIN RESULTS OBTAINED

3.RESULTS

For both SLP and g700, after the Scree test we retained the first seven principal components. These explain 88,6% and 89,9% respectively, from the total variance (fig.2) .

Scree test for the SLP



Scree test for g700

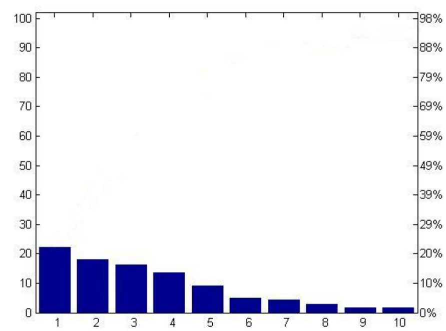


Figure 2. Scree test for the selection of the principal components for slp and g700

In order to see the most significant spatial variability patterns associated to these first seven PCs we represented the (first seven) EOFs associated to them (fig.3).

We can see here that NAO-like pattern is represented by the first EOF of the SLP and that there is a clear correspondence between this first EOF in SLP and the second in the g700 Also, there is a good spatial correspondence in the representation of the second EOF in SLP and the first EOF in g700 and in all the following components (we are not so much interested in the sign as in the spatial structure) .

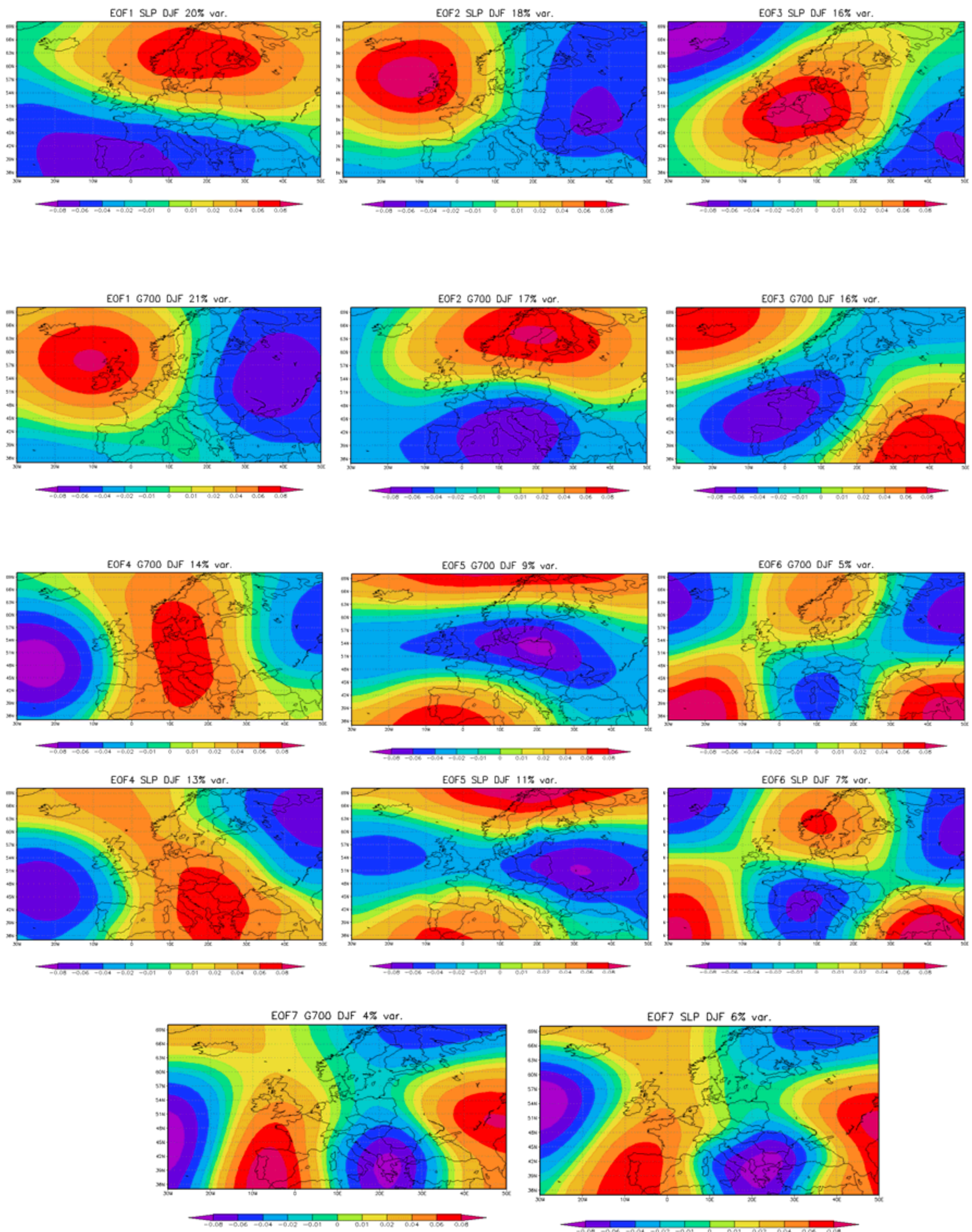


Figure 3. The first seven EOFs for SLP and g700 respectively

For the selection of the cluster number and the centroids we used the variation patterns established by the PCs retained. For each principal component we selected the values that are either larger than 1 or lower than -1 and have the rest of the PCs between these values. For the principal components for the SLP we can see (fig.4) that we can take a maximum number of 13 components (component of PC2 , negative phase has almost no days associated) .

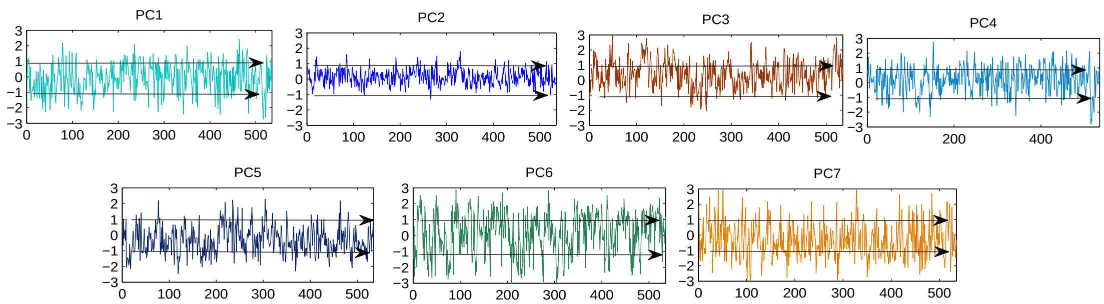


Figure 4. Selection of Cluster number and centroids for the slp

For the principal components for the g700 we selected a smaller number of cases because as we can see (fig.5) there are some PCs where we have no value above 1 or under -1 so that we could construct a cluster and a group. We decided, for the g700 to consider an initial number of clusters of 10

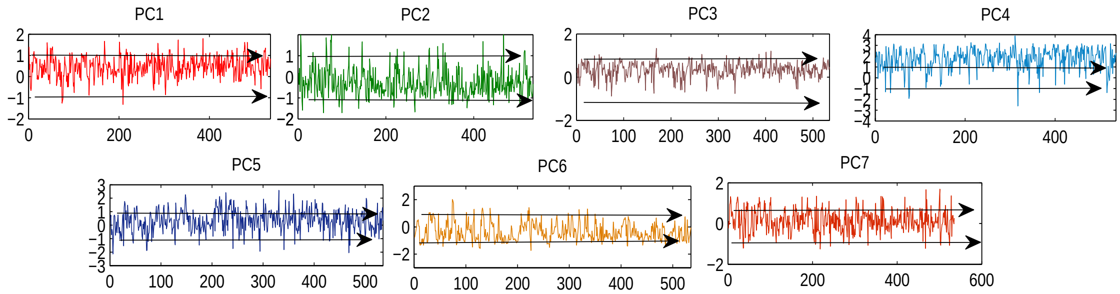


Figure 5. Selection of Cluster number and centroids for the g700

As we mentioned before, from analyzing the positive and negative phase of the PCs for both SLP and g700, we selected the centroids and the number of clusters.

After selecting the number of centroids and the number of clusters, first we performed a cluster analysis on the SLP field and for each cluster obtained we associated the corresponding g700 values. As an additional study – we performed , independently, the k-means on the g700 field .

For the independent analysis for SLP and g700, we decided to consider a number of 13 clusters for the SLP and a number of 10 for the g700. Afterwards, based on the analysis of these results (the clusters obtained) we did a regrouping of clusters into a smaller number of groups to obtain the most representative circulation patterns. The regrouping was made based on an empirical analysis of the similarity of the meteorological patterns obtained.

After the regrouping of the clusters we obtained 7 clusters for the SLP and 5 clusters for g700, we realized composite maps, for each cluster (averaging the values for SLP/g700) and for each SLP cluster we associated the g700 field.

Table1.SLP Clusters

Cluster1	Cluster 2	Cluster3	Cluster4	Cluster5	Cluster6	Cluster7
Zonal	NW	High lat zonal	SW	NW(low press. undef. flow)	E	SE
204 days	34 days	19 days	54 days	70 days	78 days	76 days

From the analysis of the clusters for SLP we can notice that the greatest number of days (Clusrter 1) is associated to a zonal flow over Romania (Table1 and Fig.6).

This zonal circulation is the most frequent situation associated to heavy precipitation over Romania during winter and it is a high possibility that this brings more rainy precipitation then snow considering the fact that will bring humidity from the Atlantic Ocean.

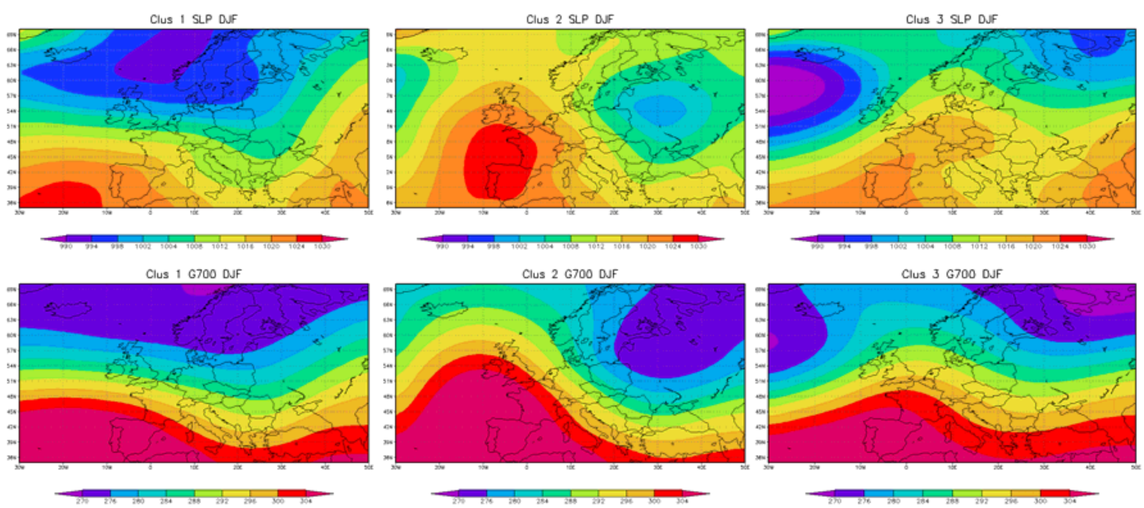


Figure6. First three clusters for SLP and their g700 associated clusters

This situations are followed by E and SE circulation patterns (cluster 6 and 7) . This are typical situations of heavy snow storms over Romania and the Siberian high might be the principal factor in these situations. A significant number of days (70-Cluster 5) is included in a NW circulation pattern which at Romanian level could be characterized as low pressure undefined flow .

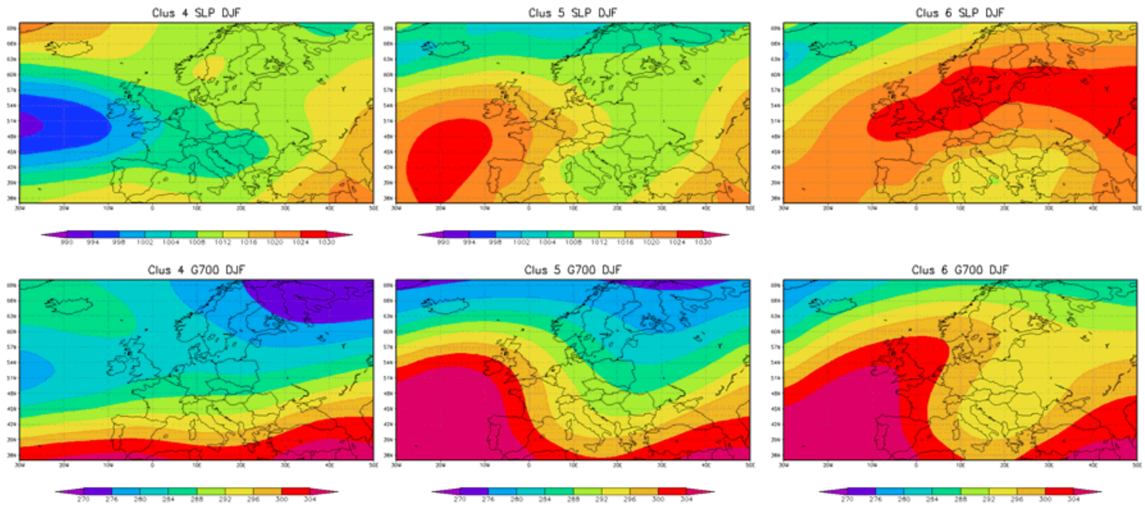


Figure7.The following three clusters for SLP and their g700 associated clusters

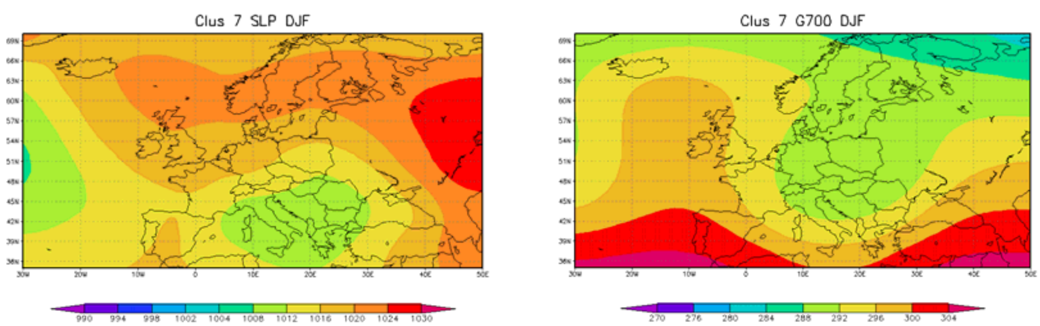


Figure8. Cluster 7 for SLP and his associated cluster 7 for g700

Also, in Cluster 7 we can observe a blocking situation at European level that is not usually associated to heavy precipitation. Still, though at European level we have this blocking situation, we can see that over Romania we have a E, NE circulation that is also sustained by the g700 pattern where it is clear that we have a circulation in altitude that could bring humid air.

We also performed an individual clustering of the g700 field and we obtained the main circulation pattern associated to heavy precipitation days (the biggest number of days) as being also a zonal type circulation (see Table2 and figure 9)

Table2. G700 Clusters

Cluster1	Cluster 2	Cluster3	Cluster4	Cluster5
S	Zonal	SW	NW (high latitude zonal flow)	S
75 days	227days	44 days	175 days	14 days

On the second place this circulation pattern is followed by a NW circulation, a high latitude zonal flow .

We can notice here that even the analysis was made separately, we have an agreement between the most significant pattern for SLP associated to heavy precipitation and the most significant one in the g700, namely, they are both zonal flows bringing humid air from the Atlantic Ocean.

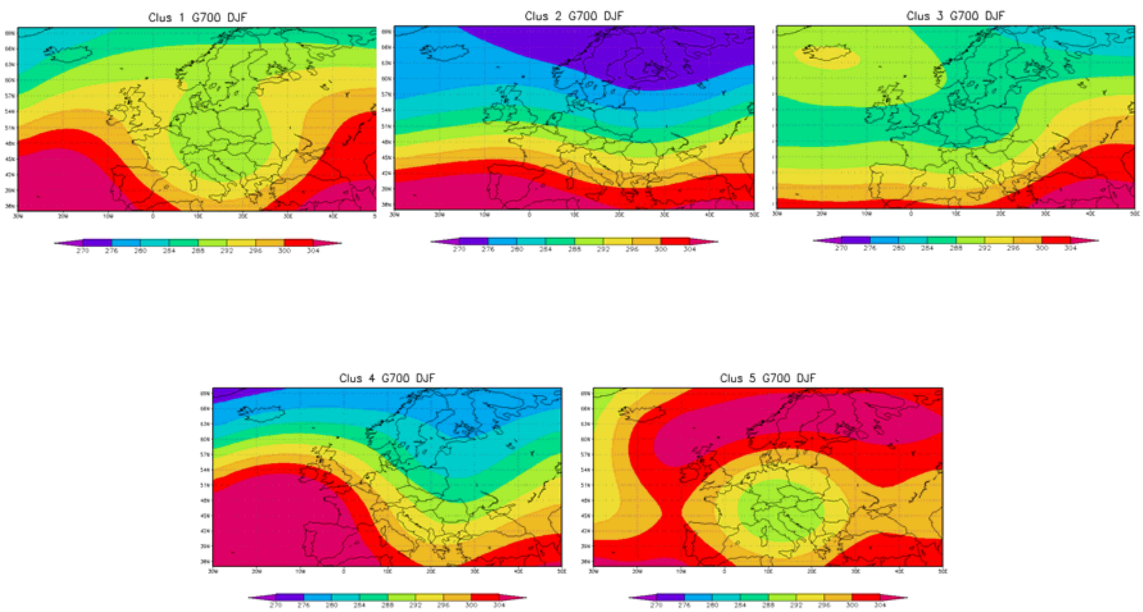


Figure9. Clusters for g700

4.SUMMARY AND FURTHER WORK

We realized the classification of circulation patterns associated to heavy precipitation days at 30 stations in the intra Carpathian region of Romania using the principal components analysis and k-means clustering method.

The classification was made both in association and independently for SLP and g700 fields. After the k-means classification we got 7 and 5 respectively clusters. For having a better understanding of both the sea level pressure and high altitude patterns associated to heavy precipitations we associated to the seven clusters we have obtained for SPL their correspondents in altitude for g700 field.

The central result from analyzing the clusters in association or separately showed the same – that the most frequent circulation patterns associated to heavy precipitation over Romania, during cold season (December, January , February) are related to a zonal circulation that brings humidity from the Atlantic Ocean. The following pattern as importance, that brings heavy snows over Romania is related to air flows coming from Siberia – S and SE type circulations that are responsible for heavy snow falls over Romanian territory in the winter months.

As a more ‘nonclasical ‘situation we got a cluster in which showed a blocking pattern at European level that could hardly be associated to extreme precipitation events over the Romanian territory.

For a better understanding of the patterns associated to extreme precipitation events and a more detailed study we plan in the future to extend this work and, as a following stage we’ll consider all the seasons and both intracharpatian – extra Carpathian regions of Romania .

5. PROJECTED PUBLICATIONS /ARTICLES RESULTING OR TO RESULT FROM THE STSM

Oral presentation in the EGU 2008 European Geosciences Union General Assembly 2008 - "*Classification of atmospheric circulation patterns associated to heavy precipitation in Romania*", Colfescu, I.; Boroneant, C.; Esteban, P.; Martin-Vide, J. in the session NH1.2 - Extreme Events Induced by Weather and Climate Change: Evaluation, Forecasting and Proactive Planning

6. CONFIRMATION BY THE HOST INSTITUTE OF THE SUCCESFULL EXECCUTION OF THE MISSION

Ioana Colfescu learned during the STMS the methods of multivariate analysis we used and assimilated their applications to our case of study (the heavy precipitation over Romanian territory) showing all the time an excellent dedication and interest in the study. Through this process we acquired an interesting scientific collaboration with the Romanian group - Constanta Boroneant and Ioana Colfescu - and ours in Barcelona - Javier Martin-Vide and Pere Esteban .

7. COMMENTS

I would like to thank to the chair of the whole management committee of the COST 733 Action and to the chair of the Action - Dr. Ole Einar Tveito - for offering me the opportunity of a STSM at the University of Barcelona, Group of Climatology.
I would like to thank to my colleague and advisor - Dr. Constanta Boroneant - who worked, advised, and helped me in all this study during all the STSM .
A big 'thank you' to the Spanish colleagues Pere Esteban and Javier Martin-Vide for sharing their experience and time with us through all the short term scientific

mission period and afterwards. For me as being somehow new in this field was the first experience of this kind and I think I was lucky to have suck great collaborators as Profesor Vide and Pere.

Also thanks to my colleagues from the National Meteorological Administration Ortansa Jude and Laura Dima - forecaster meteorologists from the National Forecasting Center, National Meteorological Administration, Bucharest, Romania for their comments and helpful discussion to interpret the results.

REFERENCES

Esteban, P., Jones, PD, Martin-Vide, J., and Mases, M. 2005: Atmospheric Circulation Patterns Related To Heavy Snowfall Days In Andorra, Pyrenees. *Int. J. Climatol.* **25**: 319–329.

Esteban, P., Martin-Vide, J., and Mases, M. 2006: Daily atmospheric circulation catalogue for western Europe using multivariate techniques. *Int. J. Climatol.* **26:** 1501-1515

Huth R. 1996. An intercomparison of computer-assisted circulation classification methods. *Int. J. Climatol.***16:** 893-922.

Romero R, Sumner G, Ramis C, Genov´es A. 1999. A classification of the atmospheric circulation patterns producing significant daily rainfall in the Spanish Mediterranean area. *International Journal of Climatology* **19**: 765–785.

Fitzharris, B.B. 1987. A climatology of major avalanche winters in western Canada. *Atmos.- Ocean*, **25**, 115-136.