Project no. GOCE-CT-2003-505539

Project acronym: ENSEMBLES

Project title: ENSEMBLE-based Predictions of Climate Changes and their Impacts

Instrument: Integrated Project

Thematic Priority: Global Change and Ecosystems

Deliverable D2B.5: Methodology for Markov chain modelling of sequences of atmospheric circulation patterns for implementation with a conditional model of extreme hydro-meteorological events

Due date of deliverable: February 2006
Actual submission date: Version 1 - March 2006; Version 2 - June 2006

Start date of project: 1 September 2004 Duration: 60 Months

Organisation name of lead contractor for this deliverable: NIHWM

Revision [2]

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	x
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the Consortium (including the Commission Services)	

ENSEMBLES RT2B

Deliverable D2B.5

Methodology for Markov chain modeling of sequences of atmospheric circulation patterns for implementation with a conditional model of extreme hydro-meteorological events

Version 2 : 14 June 2006

C. Mares, Ileana Mares and Antoaneta Stanciu

C. Mares
National Institute of Hydrology and Water Management, 13686 Bucharest, Romania
constantin_mares_ro@yahoo.com

Table of contents

1. Introduction

This study aims at finding optimal fitting and modelling of the geopotential and pressure fields over the Atlantic-European region, which can be considered as a signal for the fluctuations of maximum discharge level of the Danube lower basin.

The results obtained in this study on the basis of observational data will be the starting point of subsequent developments taking into account the outputs of the ensemble prediction system within the ENSEMBLES project.

The following objectives have been achieved:

a) A Markov model was applied to the significant principal components of the MEOF of the atmospheric circulation over Atlantic European region;
b) The link between the fields at large and local scales was tested;
c) Extreme Value Analysis in the Lower Danube Basin discharge time series
in the 20 th century taking into account as covariates NAO and low frequency components of atmospheric circulation over Atlantic European region.

Data consist in:

Observed monthly fields (1958-2001, source: **ERA-40)** for Atlantic European region (30^0-65^0N; 50^0W-40^0E):

- Sea level pressure (SLP)
- 500-1000 hPa thickness
- 500 hPa geopotential

Sea Surface Temperature Anomalies (SSTA) from tropical Atlantic (22.5 ^{0}N-22.5^{0}S; 67.5^0W-12.5^0E). Source: *Hadley Centre*

Observed field at local scale:

- Daily and monthly values of the discharge level of the Danube lower basin (Orsova) in the period 1900-2005.
- Monthly mean temperature at Bucharest – Filaret (1958-2001)

Methods

- Decomposition in Multivariate Empirical Orthogonal Functions (MEOF)
- Markov models applied to the significant principal components of the MEOF
- Extreme Value Theory (EVT)
- Teleconnections between local scale fields and atmospheric circulation at large scale (Linear multiple regression).

2. Analysis by MEOF of the atmospheric circulation over Atlantic-European region and modelling by Markov process

An analysis by MEOF of three monthly atmospheric fields has been achieved for all months - namely 528 values of the period 1958-2001.

Figures 1-3 show the modes 1, 2, and 3 of the MEOF decomposition of sea level pressure, 500-1000 hPa thickness and 500 hPa geopotential fields. Figures 1 abc, 2 abc and 3abc are the spatial distribution of mode 1, 2 and 3 respectively, while Figures 1d, 2d, and 3d show the temporal components of the first MEOFs. In Figures 1d, 2d, and 3d the red and green lines represent the 90th and 10th percentiles respectively.

The first three MEOF components explain 46% of the total variance of the analysed fields, 18%, 15% and 13% respectively.

Before modelling by a Markov process the atmospheric circulation filtered through MEOF, we should decide how many components we keep. After testing, it was concluded that satisfactory results for the behaviour of the time series of Danube lower maximum discharge, depending on the MEOF PCs, could be obtained keeping only three components.

The atmospheric circulation modelling by Markov process has been carried out using two methodologies.

The first of them consists in constructing a Markov chain of the first order with four states, for each of the first PCs. The four states have been obtained considering PCs values associated with Danube lower basin discharge values:

The transition probability matrix for the first PC is the following:

$$M_{PC1} = \begin{pmatrix} 0.468 & 0.111 & 0.140 & 0.281 \\ 0.196 & 0.309 & 0.361 & 0.134 \\ 0.174 & 0.236 & 0.438 & 0.153 \\ 0.412 & 0.123 & 0.184 & 0.281 \end{pmatrix}$$

In Fig. A, the graph of the main transitions of the states of the first temporal component of the MEOF decomposition of atmosphere circulation over Europe, associated with Danube lower basin discharge at Orsova (DDO) is presented. The four states are defined as:

S1: $PC1 \leq 0$; $DDO \leq 0$
S2: $PC1 \leq 0$; $DDO > 0$
S3: $PC1 > 0$; $DDO > 0$
S4: $PC1 > 0$; $DDO \leq 0$

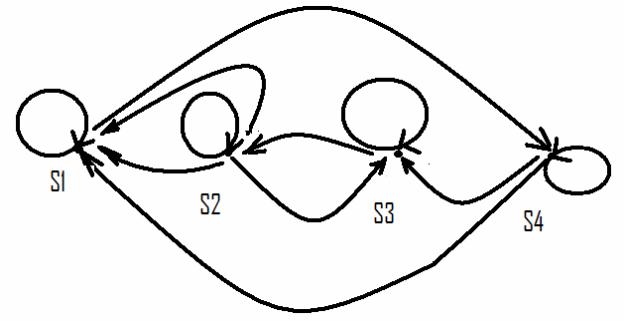


Fig. A. Graph of Markov chain with four states for PC1 of atmosphere circulation over Europe, associated with Danube lower basin discharge at Orsova (DDO).

The second methodology consists in constructing a Markov model in the multivariate EOF space of three atmospheric fields in the period 1958-2001.

Among the recent applications of this approach, those of *Xue et al*. (2000) for prediction of ENSO events and those accomplished by *Chen and Yuan* (2004) for the simulation and forecast of short-range climate changes in the Antarctic can be mentioned.

According to *Chen and Yuan* (2004), we can decompose a matrix **V** into MEOFs (spatial patterns) **E** and their corresponding PCs (time series) **P**:

$$\boldsymbol{V=E\ P^T} \qquad (1)$$

where the columns of **E** are orthogonal and the columns of **P** are orthonormal: the superscript T denotes matrix transposition.

The space reduction is accomplished by (1) for the first few modes.

The Markov model is calculated using a single step correlation matrix, that is, a transition matrix in accordance with the linear relation:

$$\boldsymbol{P_{i+1} = A\ P_i + e_i} \qquad (2)$$

where i is the i-th month and e is the model fitting error.

Multiplying (2) with P_i^T and making a mean over time, we get:

$$\left\langle P_{i+1}P_i^T \right\rangle = A\left\langle P_i P_i^T \right\rangle + \left\langle e_i P_i^T \right\rangle \qquad (3)$$

where the angle brackets indicate the average over time.

If the model had the best fitting, there would not be any correlation between e_i and $\mathbf{P_i^T}$ so that **A** could be calculated thus:

$$A = \left\langle P_{i+1}P_i^T \right\rangle \left\langle P_i P_i^T \right\rangle^{-1} \qquad (4)$$

In our case **A** is calculated regardless of the season.

We have calculated matrix A both for the first three and for the first 10 PCs of MEOF decomposition for the period 1958-2001.

For the three component matrix A is:

$$A = \begin{pmatrix} 0.123 & -0.025 & -0.063 \\ 0.043 & 0.085 & 0.044 \\ 0.065 & 0.008 & 0.067 \end{pmatrix}$$

The verifications for the years 2001-2004 have been encouraging.

This prediction is compared to the forecasts estimated through a 24-predictor regression equation, the first 12 predictors being the principal components (PC) of the EOF representation (65% of the variance) of the tropical Atlantic SSTA with a lag from 1 to 12 months with respect to the predictand, the following predictors being just the principal components of the three atmospheric fields, considered with the same lag with respect to the dependent variable. From the analysis of outputs, it was concluded that the six-month previous SST PC is a good predictor together with the one month previous MEOF PC, for the atmospheric circulation over Europe.

3. Teleconnections between the Atlantic-European atmospheric fields and local scale variables

These teleconnections have been idenitified between monthly mean values, using as predictands the discharge level of the Danube lower basin (Orsova) and Bucharest - Filaret temperatures, and as predictors the first 10 temporal MEOF components or the time series from each grid point where the atmospheric fields are defined. Simultaneous connections as well as lag connections for the 1958-2001 period have been tested.

For the series of monthly discharges, in the case of the 10 predictors, using a multiple linear regression, the most significant results from the statistical point of view, have been obtained in the case of one-month lag, meaning that the atmospheric circulation over Europe at moment *t* is a good predictor for the Danube lower basin discharge at moment *t+1*.

This is expected, as the Danube discharge level when it enters Romania is an integrator of the central Europe precipitation.

For exemplification, we give here the spatial distribution of the correlation between the monthly August mean of SLP, 500 - 1000 hPa thickness and 500 hPa geopotential, and the Orsova discharge level in September (Fig. 4a, b, c). These two months have been chosen because in September 2003 occurred the second minimum Orsova discharge level, the absolute minimum being that of 1947 (Fig.5).

The 2003 heat waves that affected many parts of Europe were investigated in *Beniston and Diaz* (2004) and *Beniston and Stephenson* (2004). These heat waves were an important factor, which led to the very lower discharge level in the Danube basin.

Another example of the signal of atmospheric circulation in Danube discharge level, with a one-month lag, is presented in Fig. 6, where the April time series of the Danube discharges at Orsova and the 500 hPa atmospheric blocking index in March can be seen.
Figure 7 shows the spatial distribution of the coefficients of correlation between the Bucharest-Filaret monthly mean temperature and the three atmospheric fields for August. As we expect for the Bucharest-Filaret monthly mean temperatures, the 500-1000 hPa thickness signal is the highest and is simultaneous (Fig.7b).

These are many studies regarding the connections between circulation at large scale and characteristics of climate variables at regional (local) scale. Among them we mention those carried out by *Goodess and Palutikof* (1998), as well as *Goodess and Jones* (2002) and, within the STARDEX project, *Haylock and Goodess* (2004).

4. Extreme Value Analysis in the Danube lower basin discharge time series with NAO and PCs of MEOF as covariates

4.1. Daily time series

The daily discharge time series in the lower Danube basin have been considered for the 1900-2005 period. Orsova station, situated in the south-western part of Romania, was considered as representative for this analysis.

The extreme value theory (EVT) is applied for the study of daily maximum discharges incorporating some covariates. Two methods are applied for fitting the data to an extreme - value distribution: *block maxima* and *peaks over thresholds* (POT).

Using the *block maxima* approach associated with the use of the generalized extreme value (GEV) distribution, the annual and spring (March, April, May and June) maxima of daily discharge for 1900-2005 have been analysed, as well as the monthly maxima of daily discharge for the 1958-2001 period.

In Fig. 8 we can see the time series of annual maximum discharge levels obtained from daily data from the period 1900-2005.

In order to estimate the parameters, the Maximum Likelihood Estimation (MLE) method was used. From the three possible types of GEV distribution, a Weibull distribution fits very well both annual and spring maxima, and monthly maxima of the daily discharges. The GEV fit diagnostics for Orsova annual daily maximum discharges are presented in Fig.9.

GEV distribution parameters (location, scale and shape) with standard errors in brackets are as follows:

μ = 9575.44 (192.2)
σ = 1748.18 (135.8)
ξ = - 0.17 (0.07)

Estimation of the confidence limits for the return period and shape parameter is based on the profile likelihood method. In Fig.10 is presented the profile likelihood for the 100-year return level (mc/s) and shape parameter (ξ) of the GEV distribution fit for Orsova annual maximum discharge dataset.

As regards the introduction of a covariate in the GEV distribution parameters, at first we have tested the winter (December, January and February) NAO as covariate.

By means of the likelihood-ratio test statistic there was obtained a value of χ^2 a little higher than that theoretically corresponding to the 1-0.05 th quantile of a χ^2 with corresponding degrees of freedom, only in the case of introducing the covariate in the location parameter.

But instead of December, January and February NAO we have introduced December, January February and March (DJFM) obtained from (http://www.cgd.ucar.edu/cas/jhurrell/indices.html).

The best result has been got by incorporating NAO from DJFM in the location parameter of GEV for the daily maximum discharge levels during spring and early summer months (March, April, May and June).

In the Generalized Pareto Distribution (GPD) analysis, the difficult problem that appears is connected to threshold selection, that is to finding the values higher than a certain limit, the method being known as POT. A too low threshold leads to biased parameter estimates, while a too high one will result in large variance of the parameter estimates.

We have three possibilities of testing the threshold:

i) Mean Residual life Plot

The result obtained by Mean Residual life Plot is shown in Fig. 11. This figure is not generally easy to interpret. The point is to find out the least threshold up to where the line is almost linear, taking also into consideration the 95% confidence bounds. It results in a threshold between 8000 and 10000 mc/s.

ii) The second method tries to find out a threshold by fitting the data several times to a GPD distribution, using a different threshold each time. By choosing 50 thresholds from 7000 mc/s to 11000 mc/s, the changes in the scale and shape parameters shown in Fig. 12 are obtained.

From these graphs we can also choose thresholds between 8000mc/s and 10000 mc/s. Above 10000 mc/s, changes occur both in the scale parameter and in the shape one.

iii) Fitting data to a Point Process Model (the graphical representation is not shown here).

After analyzing all the graphs we decided upon a POT threshold of 10000 mc/s.

GPD fit diagnostics for the Orsova annual discharge dataset with a threshold of 10000 mc/s are presented in Fig. 13, while the GPD estimated parameters together with the standard errors (brackets) are:

σ = 2154.43 (50.17)
ξ = - 0.42 (0.02)

1641 cases from 1900-2005 of discharges higher than the imposed threshold of 10 000/mc/s have been found. The 10 000 mc/s threshold is a little higher than the 95th percentile, which is 9720 mc/s. Therefore, the maximum discharges over a certain threshold are fitted with a bounded or beta distribution.

4.2. Monthly time series

We show below the results obtained for the 1958-2001 Orsova monthly maximum daily discharges. We present only the GEV analysis, as the time series is too short to also perform also a GPD study.

From fitting by GEV the following parameters resulted:

μ = 6085.08 (103.3)
σ = 2105.78 (73.9)
ξ = - 0.15 (0.03)

As in the case of annual daily maximum values the distribution is Weibull (Fig. 14).

In order to see the influence of the atmospheric circulation over the Atlantic-European region, various covariates have been tested starting by incorporating parameters μ and σ of 3 to 10

principal components (PC) of MEOF decomposition. The 10 MEOF components explain 82 % from the total variance of the fields.

As has also been found in the case of analysis by classical regression, the most significant results have been obtained for one-month lags. That is in the covariates incorporating low-frequency components of MEOF decomposition with one-month lag, regarding the discharge level of the Danube low basin (Orsova). The most significant results are obtained by incorporating the first 10 PCs of MEOF in the location parameter.

The fitting by GEV with PC1,.., PC10 as covariates is shown in Fig. 15. From the likelihood-ratio test statistic for models: M0 = gev.fit1 (without covariates) and M1 = gev.fit2 (with covariates) has resulted that: 74.9808 > 18.307 = 1 - 0.05 quantile of a Chi-square with 10 degrees of freedom and p-value = 0

We compared the results obtained by considering the first 10 PCs of MEOF decomposition as predictors for the Orsova discharges in a classical multiple linear regression with those obtained by EVT analysis incorporating the same PCs as covariates in the location parameter of a GEV distribution. The weights of PCs of MEOF decomposition of atmospheric circulation a month before the predictand monthly daily maximum of the discharge level of the Danube lower basin (Orsova) are presented in Fig.16 (CLR- Classical Linear Regression and EVT- extreme value theory).

For the first three PCs also the results are significant, but by incorporating 10 PCs in the scale parameter a slight improvement of the model occurs, while introducing three components in the scale parameter gives no improvement.

It is interesting to point out that the results obtained by incorporating PCs as covariates are comparable with the results obtained by incorporating the PC states as covariates. This fact is one of the advantages of using Markov modelling of the atmospheric circulation over the Atlantic-European region.

All the results of EVT analysis are obtained here using the extremes toolkit from an ***R*** utility **(***Gillelard and Katz*, 2005, *Katz et al*., 2005).

Extreme precipitation events and their impact upon the discharge levels of the Danube lower basin are discussed in *Mares et al.* (2005). Some of the results reported here are also presented in *Mares and Mares* (2006).

The study of the link between large-scale atmospheric circulation and local-scale extreme events is affected by the relatively short time series of geopotential fields, but this fact can be remedied by considering the daily geopotential values and also by simulation methods.

The research team from NIHWM has in view to study the occurrence of local-scale extremes associated with large-scale circulation using the procedures described here or by applying the hidden Markov models (*Bellone et al. 2000*) both for observations and also for outputs of climatic change models obtained within the ENSEMBLES project.

Acknowledgments

*The Research team thanks David Stephenson from the University of Reading for providing a part of the **R** language software.*

References

Bellone, E. J.P. Hughes and P. Guttorp, 2000: A hidden Markov model for downscaling synoptic atmospheric patterns to precipitation amounts. *Climate Research*, **15**, 1-12.

Beniston, M., and D. Stephenson, 2004: Extreme climatic events and their evolution under changing climatic conditions. *Global and Planetary Change*, **44**, 1-9.

Beniston, M., and H. F. Diaz, 2004: The 2003 heat wave as an example of summers in a greenhouse climate? Observations and climate model simulations for Basel, Switzerland. *Global and Planetary Change*, **44**, 73-81.

Chen, D. and X. Yuan, 2004: A Markov model for seasonal forecast of Antarctic sea ice**.** *Journal of Climate***, 17,** 3156-3168.

Gillelard, E. and R. W. Katz, 2005 : Tutorial for the Extremes Toolkit: Weather and Climate Applications of Extreme Value Statistics, http://www.asssessment.ucar.edu/ toolkit.

Goodess C.M. and J.P. Palutikof, 1998: Development of daily rainfall scenarios for the southeast Spain using a circulation type approach to downscaling. *International Journal of Climatology*, **18**, 1051-1081.

Goodess, C.M. and P.D. Jones, 2002: Links between circulation and changes in the characteristics of Iberian rainfall. *International Journal of Climatology*, **22**, 1593-1615.

Haylock, M. R. and C.M. Goodess, 2004: Interannual variability of European extreme winter rainfall and link with mean large-scale circulation. *International Journal of Climatology*, **24**, 750-776.

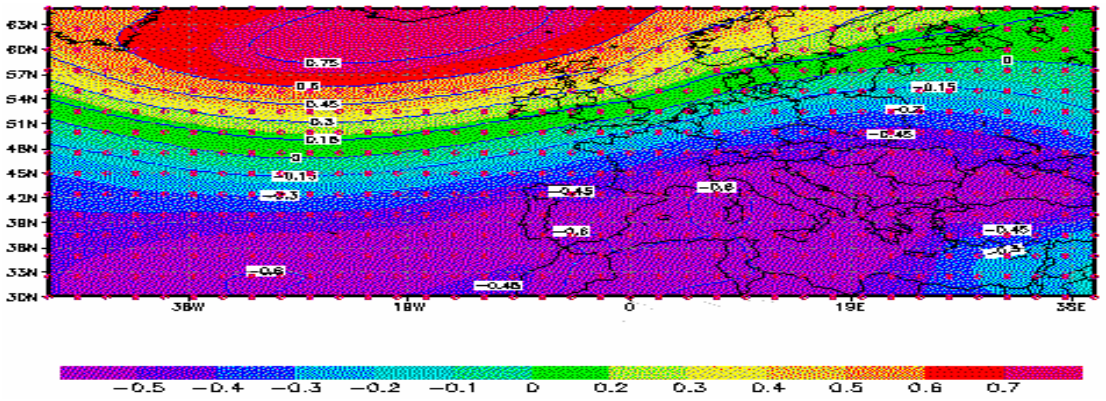
Katz, R.W., M. B., Parlange, and P. Naveau, 2002: Statistics of extremes in hydrology. *Advances in Water Resources*, **25**, 1287-1304.

Katz, R.W., S.B. Grace and M.B. Parlange, 2005: Statistics of extremes: Modeling ecological disturbances, *Ecology*, **86**, 1124-1134.

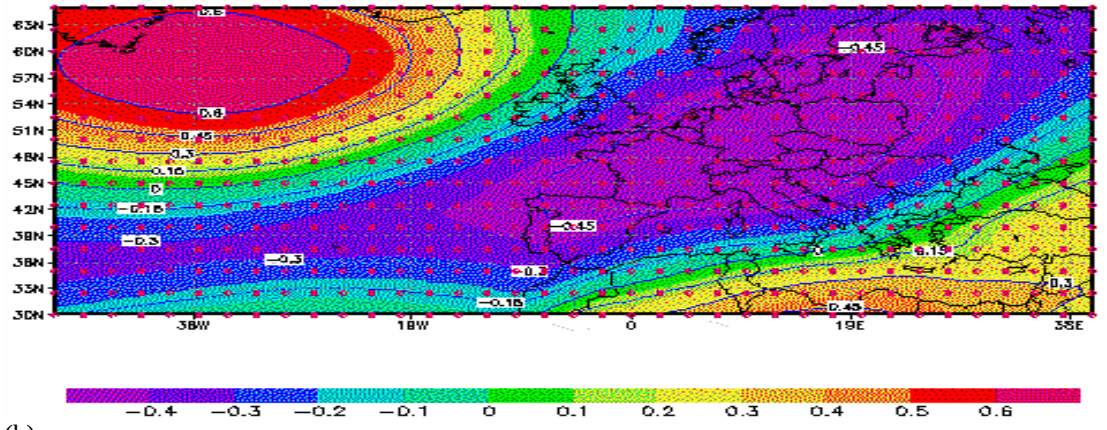
Mareş C., Ileana Mareş and Antoaneta Stanciu, 2005: Extreme climatic events in the precipitation time series in Romania and their impact in the Danube inferior basin. *European Geosciences General Assembly 2005*, Vienna, Austria, 24-29 April 2005.

Mareş C. and Ileana Mareş, 2006: *Variability and Climate Change.* Vol.1: *Statistical modelling of climate extremes* (in Romanian), Ed. University Book, 305pp.

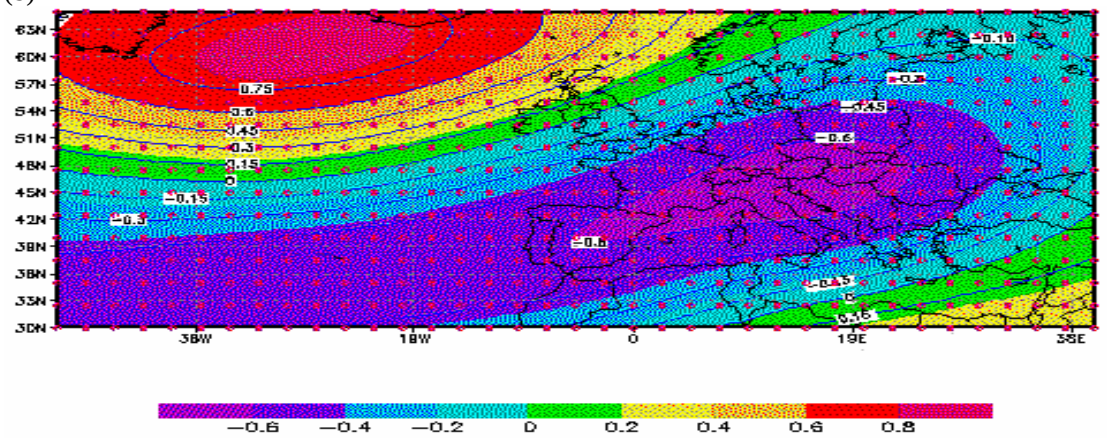
Xue, Y., A. Leetmaa and M. Ji, 2000: ENSO prediction with Markov models**.** The impact of sea level**.** *Journal of Climate***, 13,** 849-871.



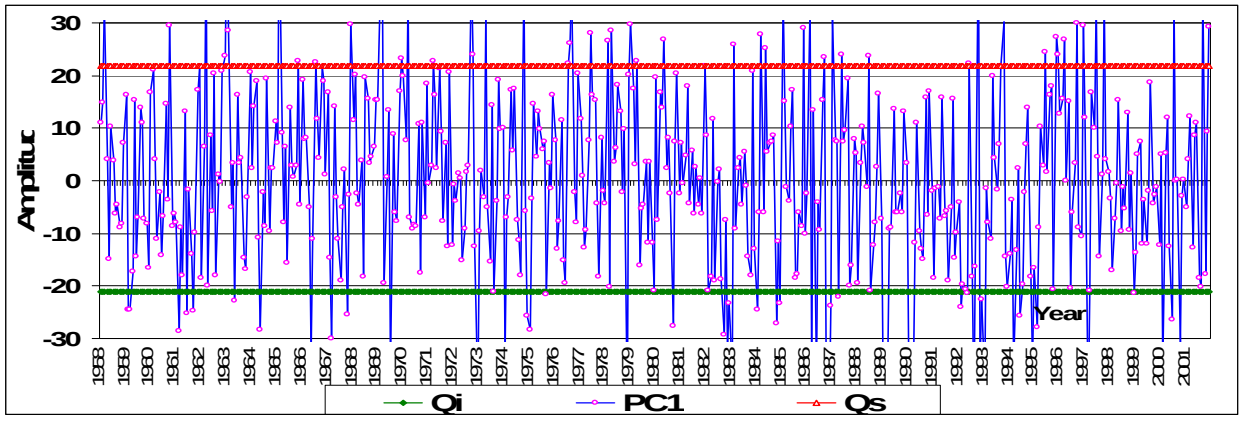
(a)



(b)

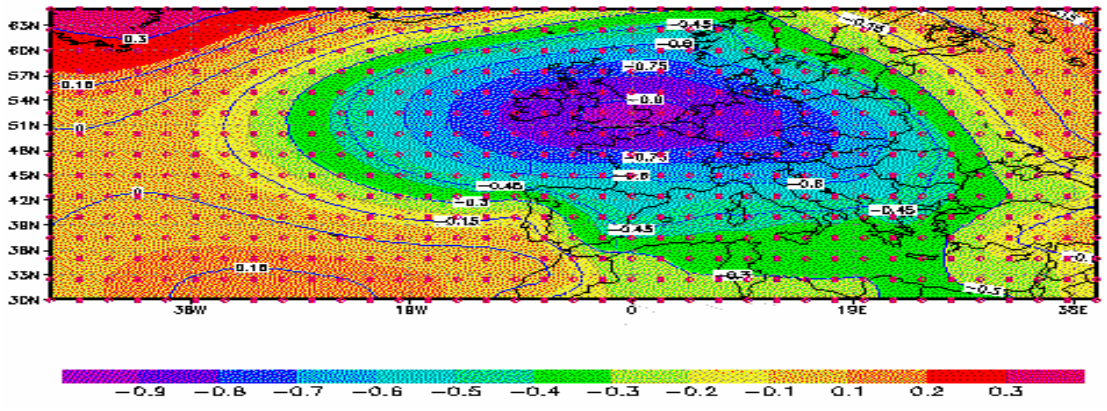


(c)

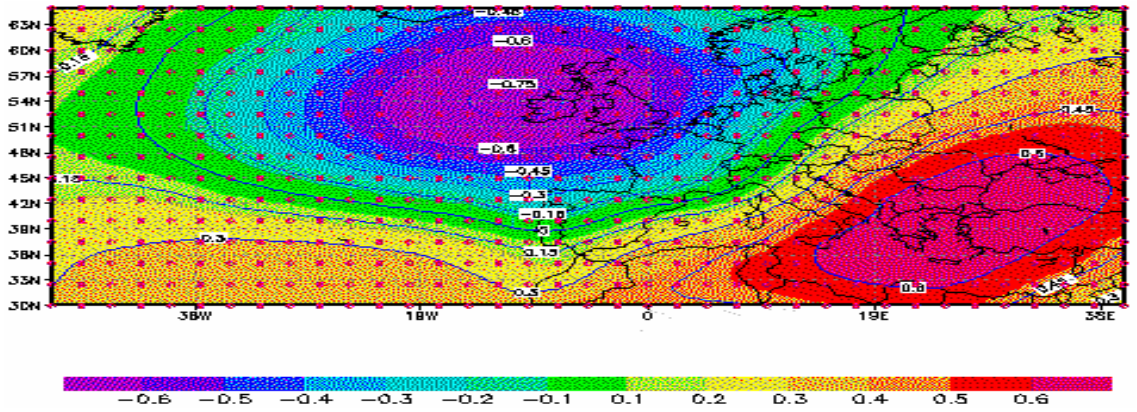


(d)

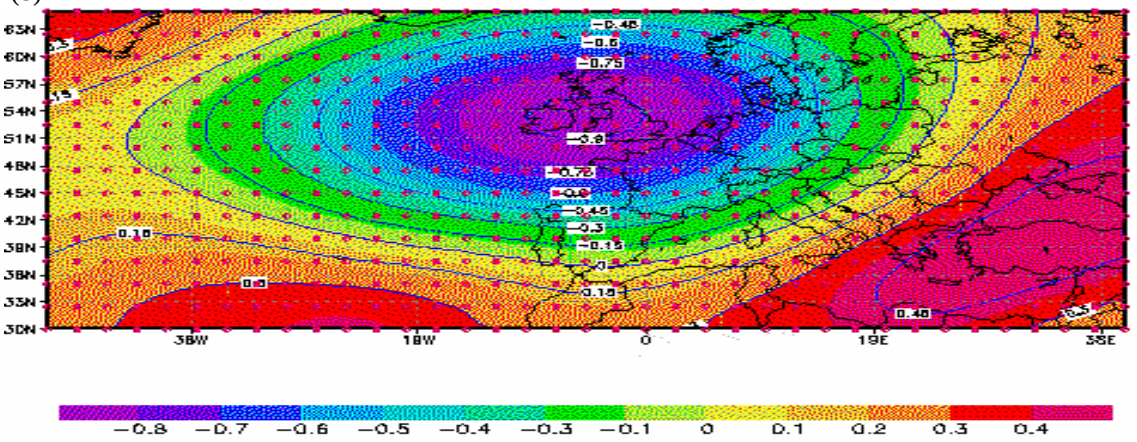
Fig. 1. (a) Sea level pressure, (b) 500-1000 hPa thickness, (c) 500 hPa of the first MEOF pattern and (d) PC1 (blue line) and the 90th (red) and 10th (green) percentiles.



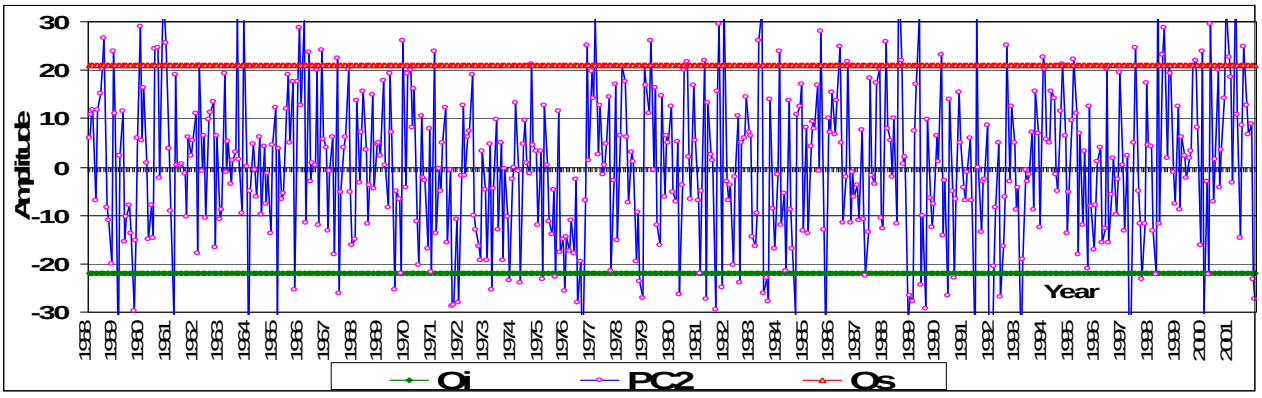
(a)



(b)

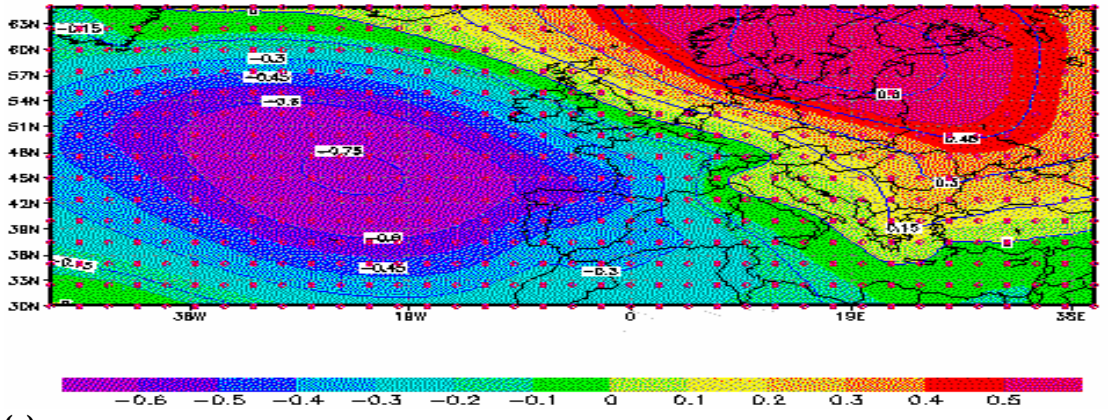


(c)

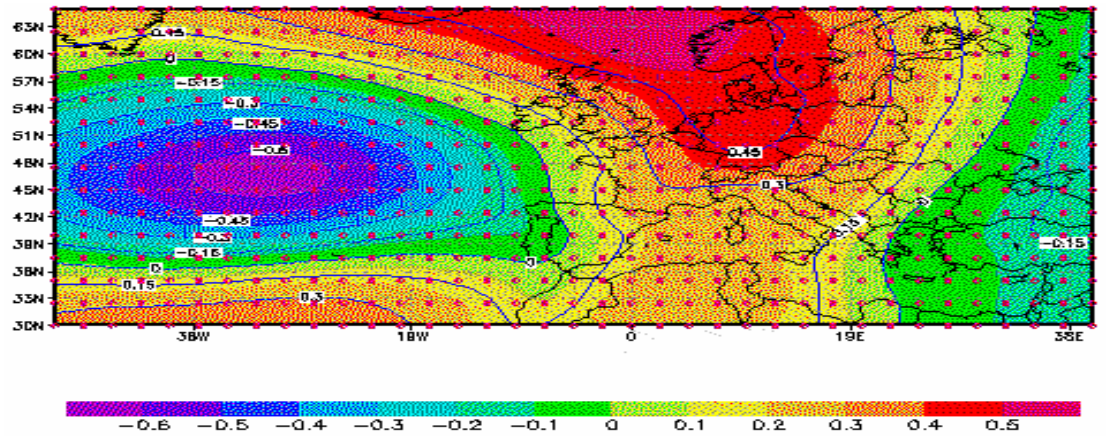


(d)

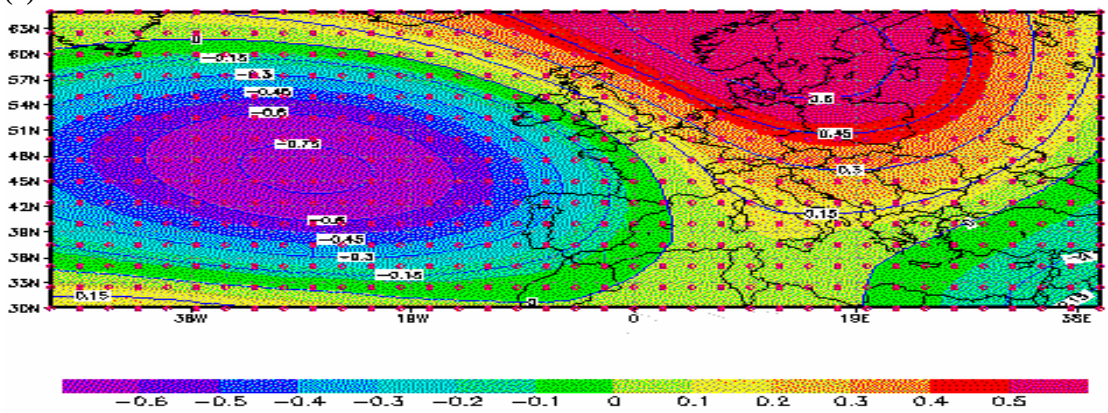
Fig. 2. (a) Sea level pressure, (b) 500-1000 hPa thickness, (c) 500 hPa of the second MEOF pattern and (d) PC2 (blue line) and the 90th (red) and 10th (green) percentiles.



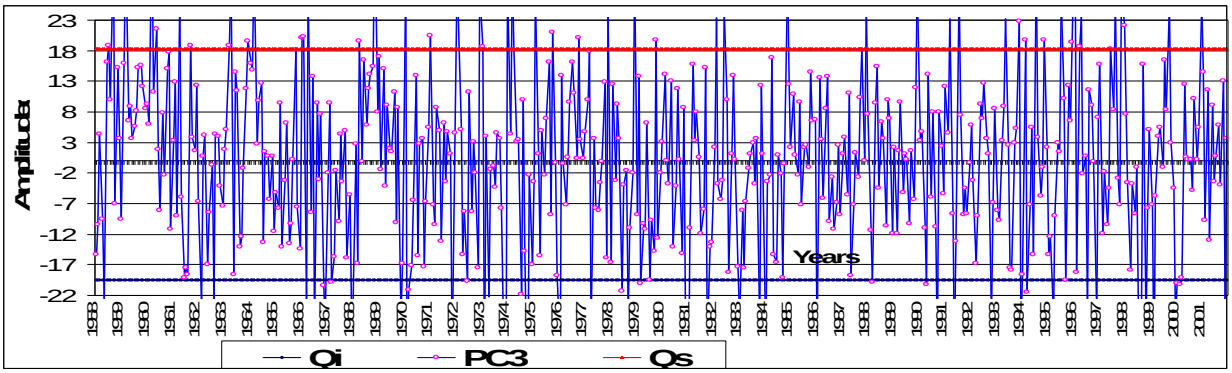
(a)



(b)

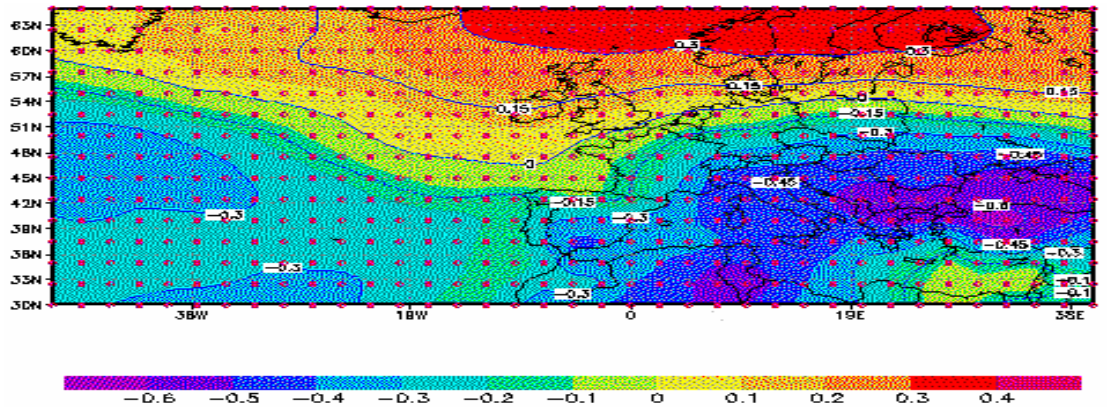


(c)

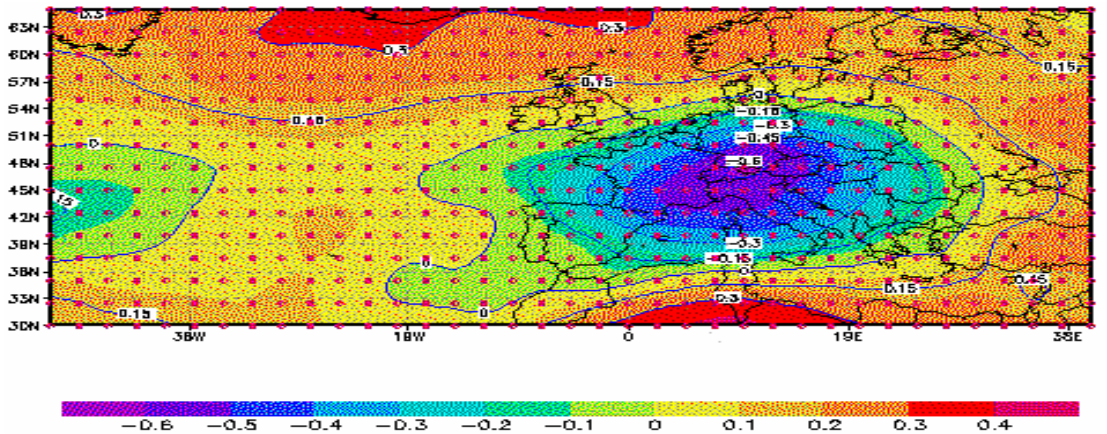


(d)

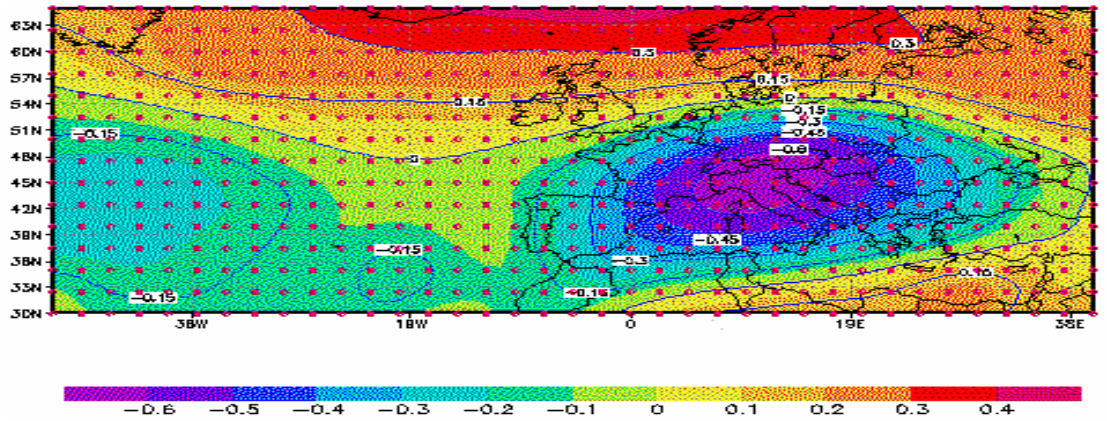
Fig. 3. (a) Sea level pressure, (b) 500-1000 hPa thickness, (c) 500 hPa of the third MEOF pattern and (d) PC3 (blue line) and the 90th (red) and 10th (blue triangle) percentiles.



(a)



(b)



(c)

Fig. 4. Teleconnection pattern of sea level pressure (a), 1000-500 hPa thickness (b), 500-hPa (c) in August and the Danube lower basin discharge (Orsova) in September (1958-2001).

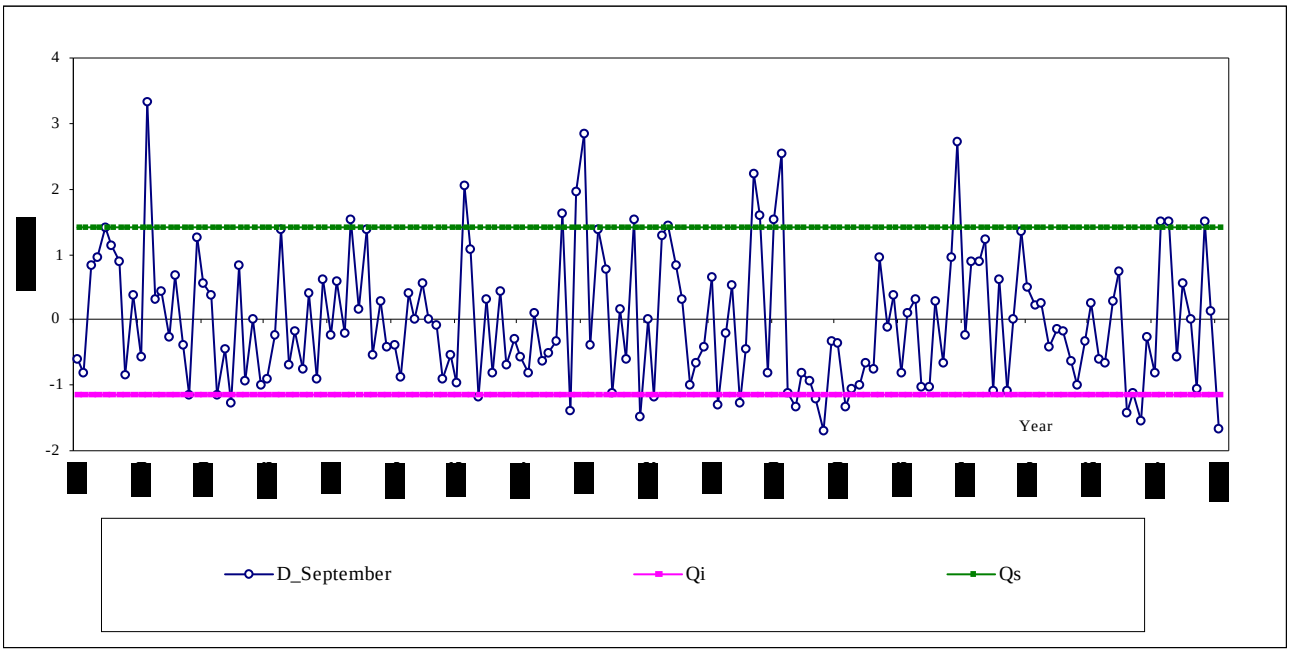


Fig. 5. Time series of discharge level (standardised) of the Danube lower basin (Orsova) for September 1840-2003 and the 90th (green line) and 10th (pink line) percentiles.

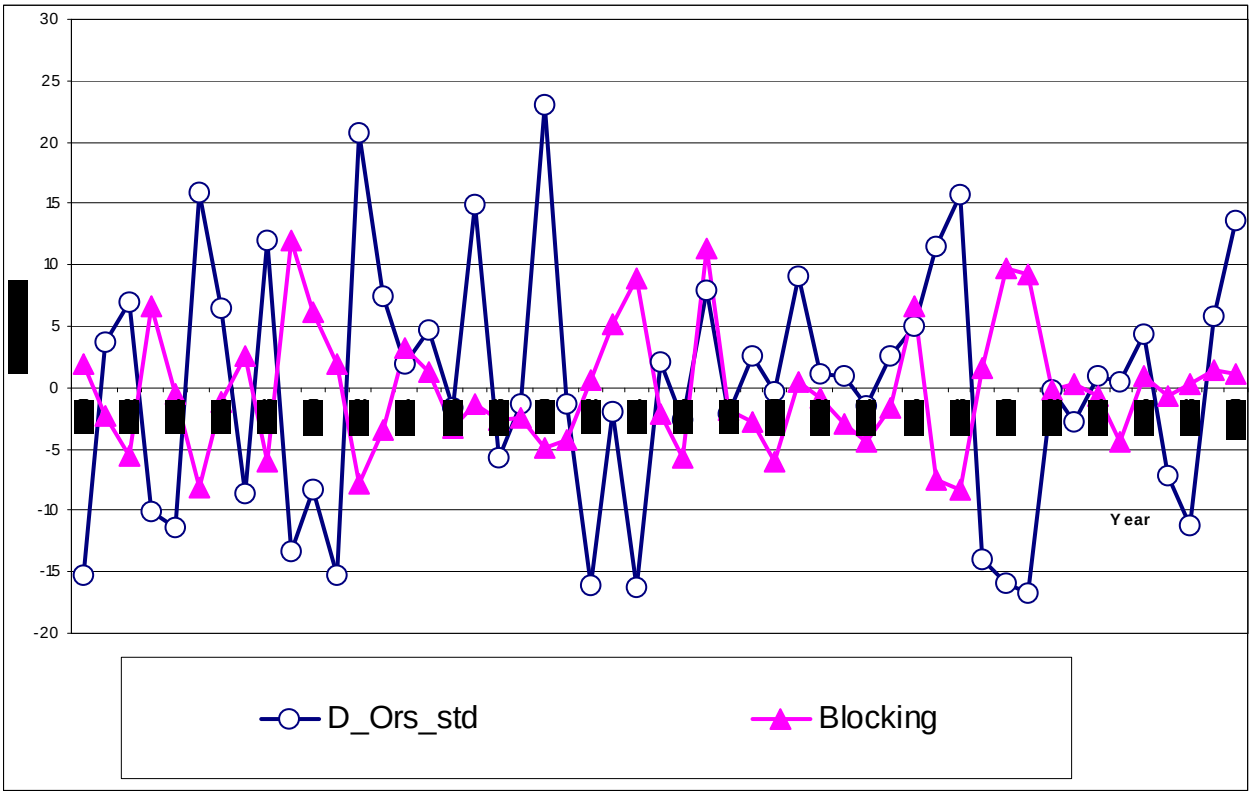
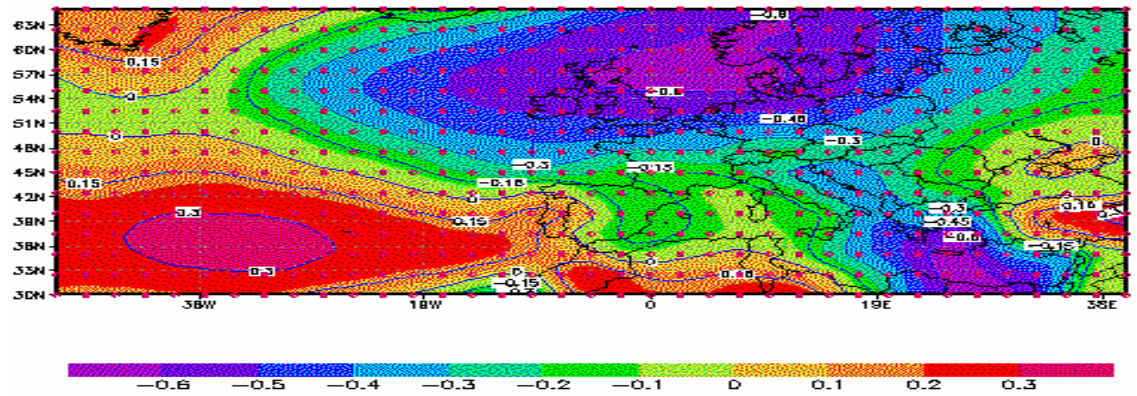
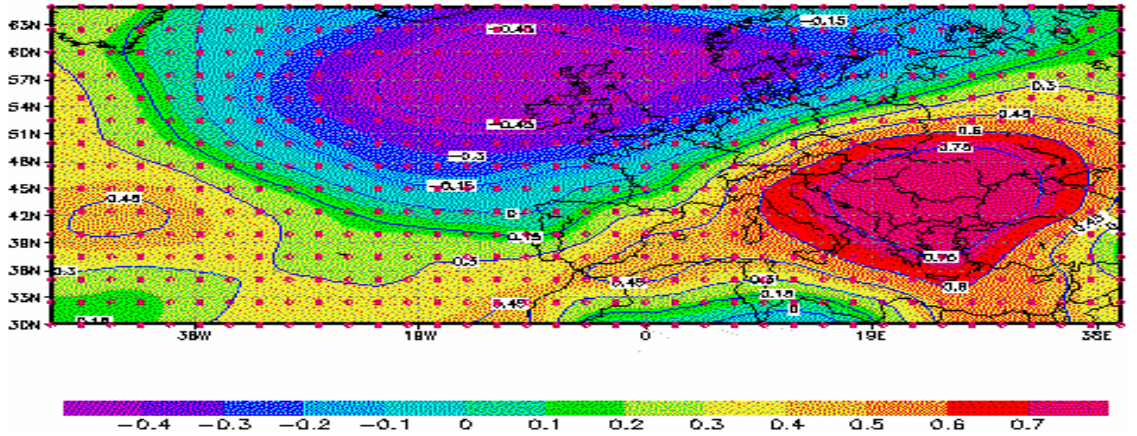


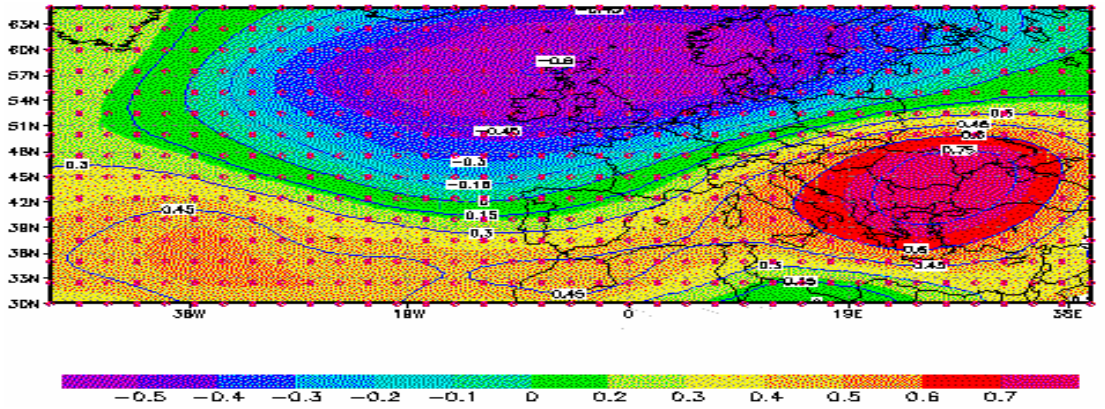
Fig. 6. Time series of discharge level of the Danube (Orsova) in April and blocking index for 500- hPa over Europe in March (r = - 0.59)



(a)



(b)



(c)

Fig. 7. Teleconnection pattern of sea level pressure (a), 1000-500 hPa thickness (b), 500 hPa (c) and Bucharest -Filaret temperature in August (1958-2001).

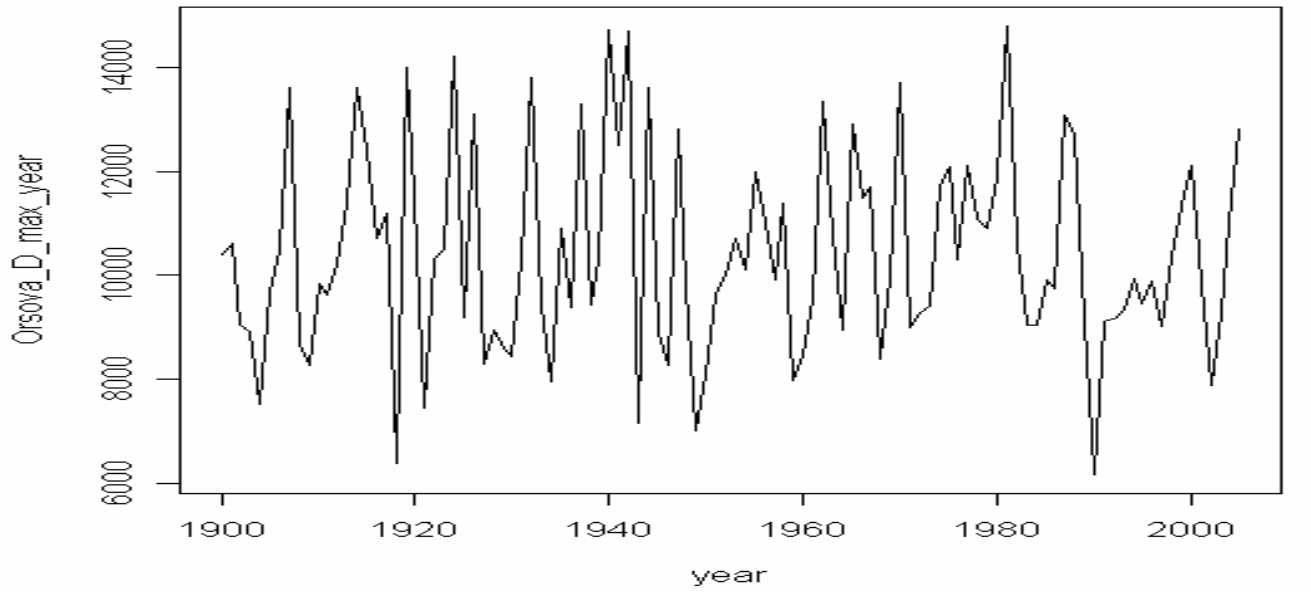


Fig. 8. Time series of annual daily maximum discharge for Danube lower basin (Orsova, 1900-2005)

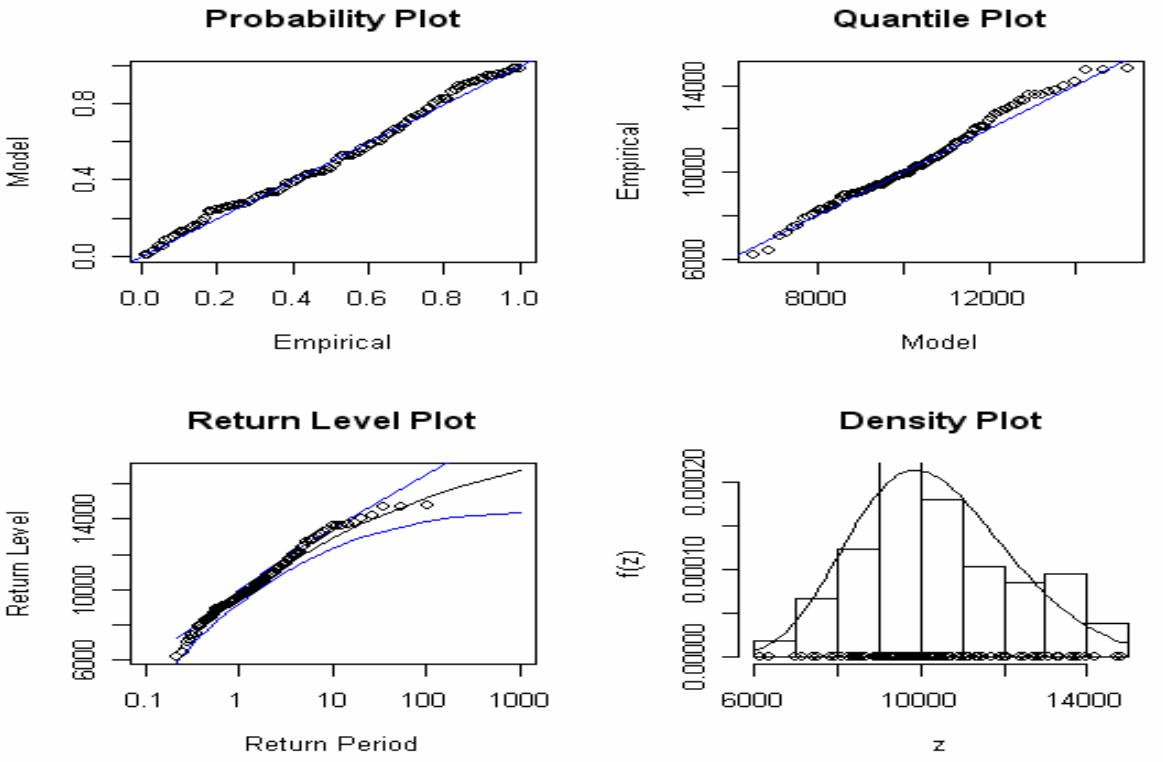


Fig. 9. GEV fit diagnostics for Orsova annual daily maximum discharge dataset. Quantile and return level plots are in mc/s.

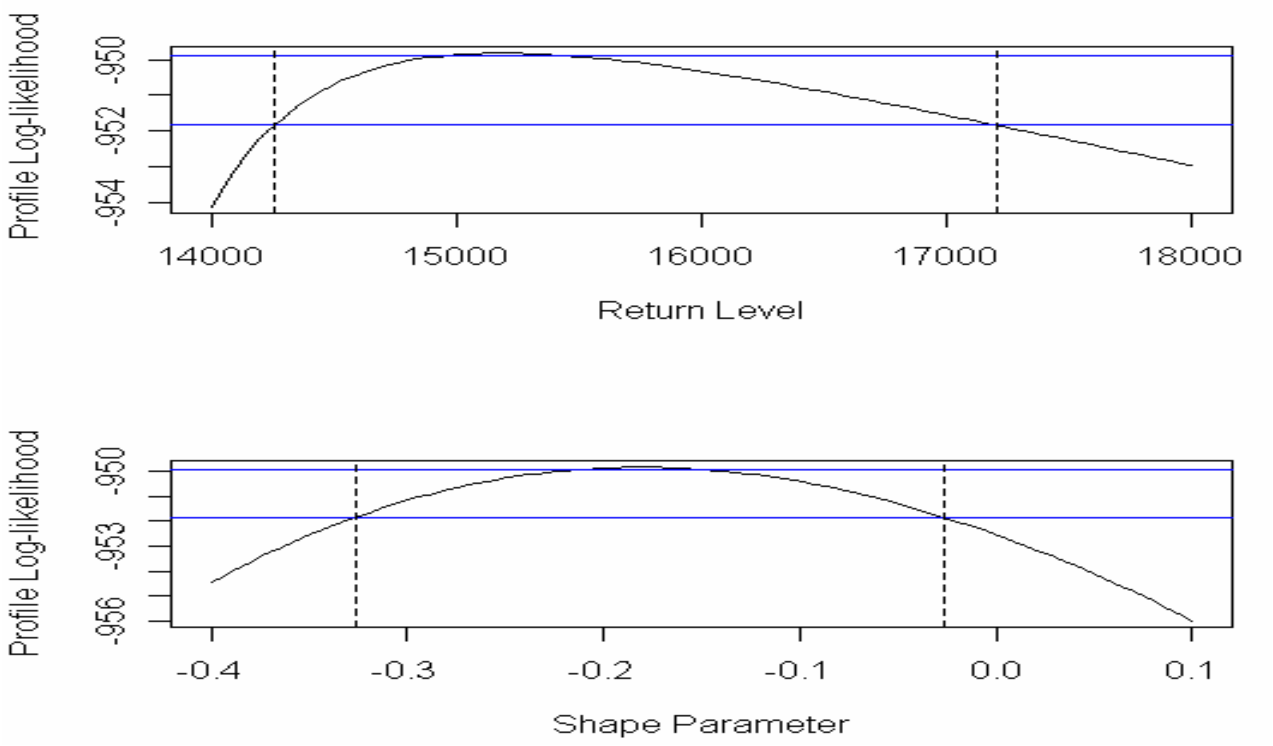


Fig. 10. Profile likelihood plots for the 100-year return level (mc/s) and shape parameter (ξ) of the GEV distribution fit for Orsova annual daily maximum discharge dataset.

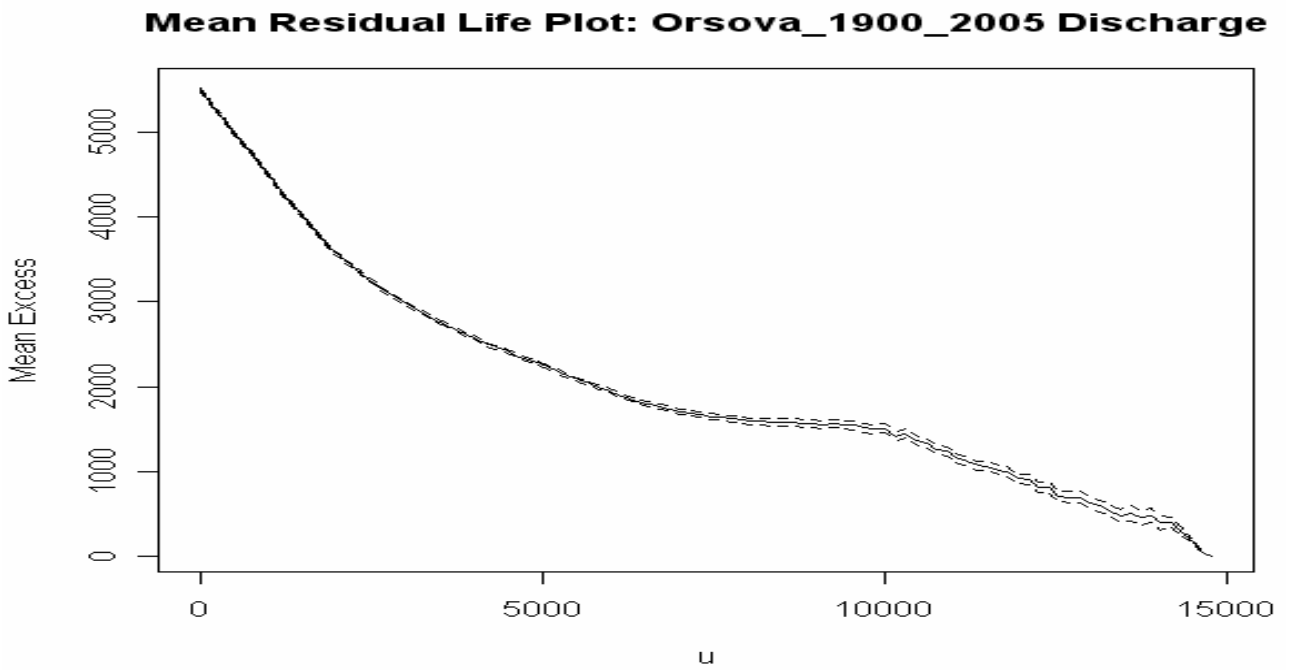


Fig. 11. Mean Residual Life Plot of Orsova daily discharge Thresholds (u) vs Mean Excess discharge (mc/s)

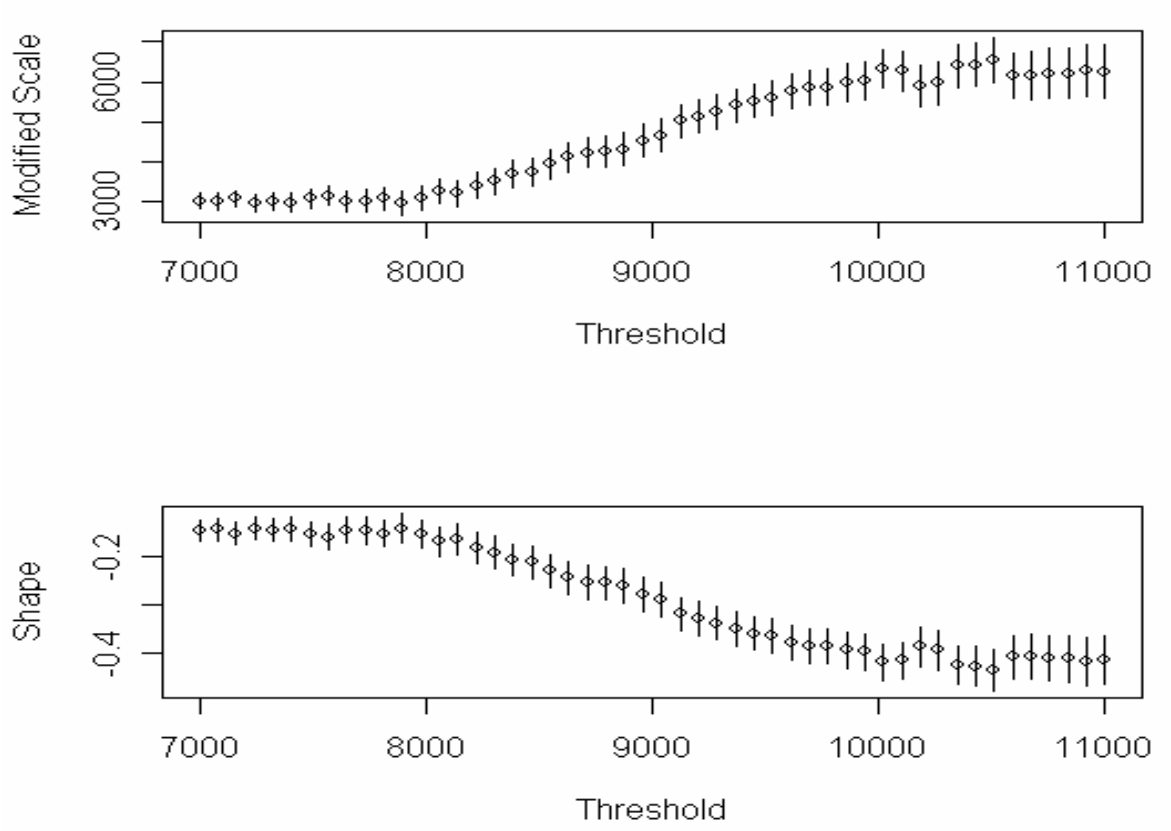


Fig.12. GPD fits for a range of 50 thresholds from 7000 mc/s to 11000 mc/s for the Orsova daily discharge.

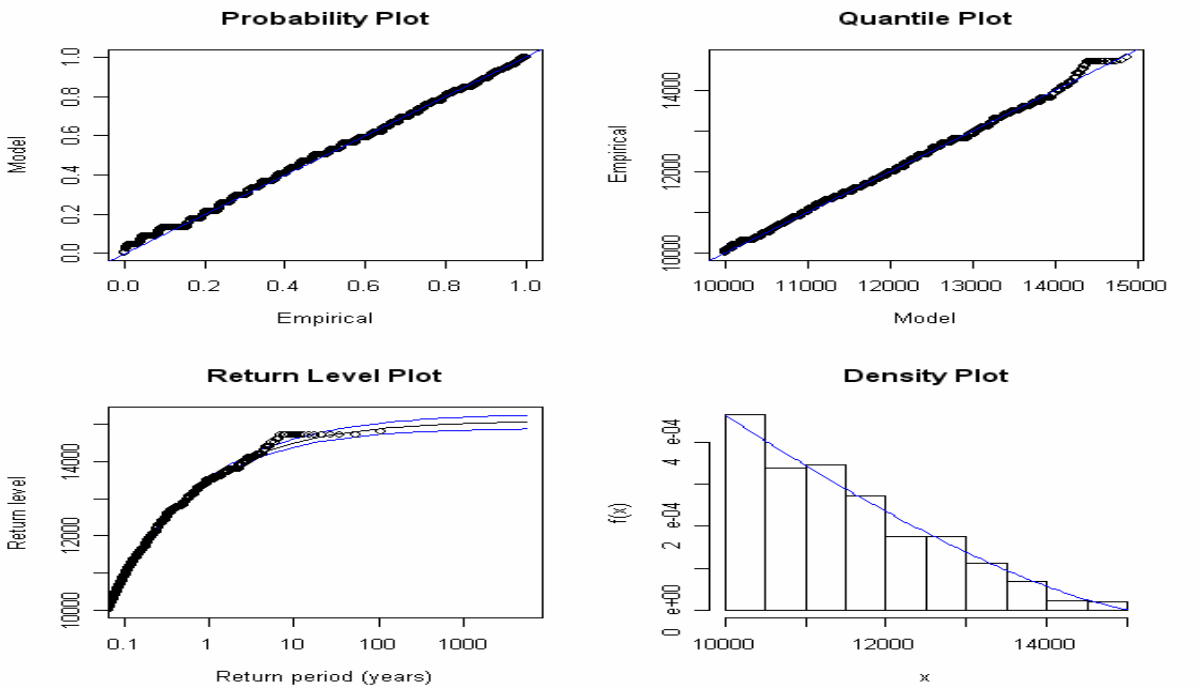


Fig.13. Diagnostic plots for the GPD fit of the Orsova daily discharge using a threshold of 10000 mc/s.

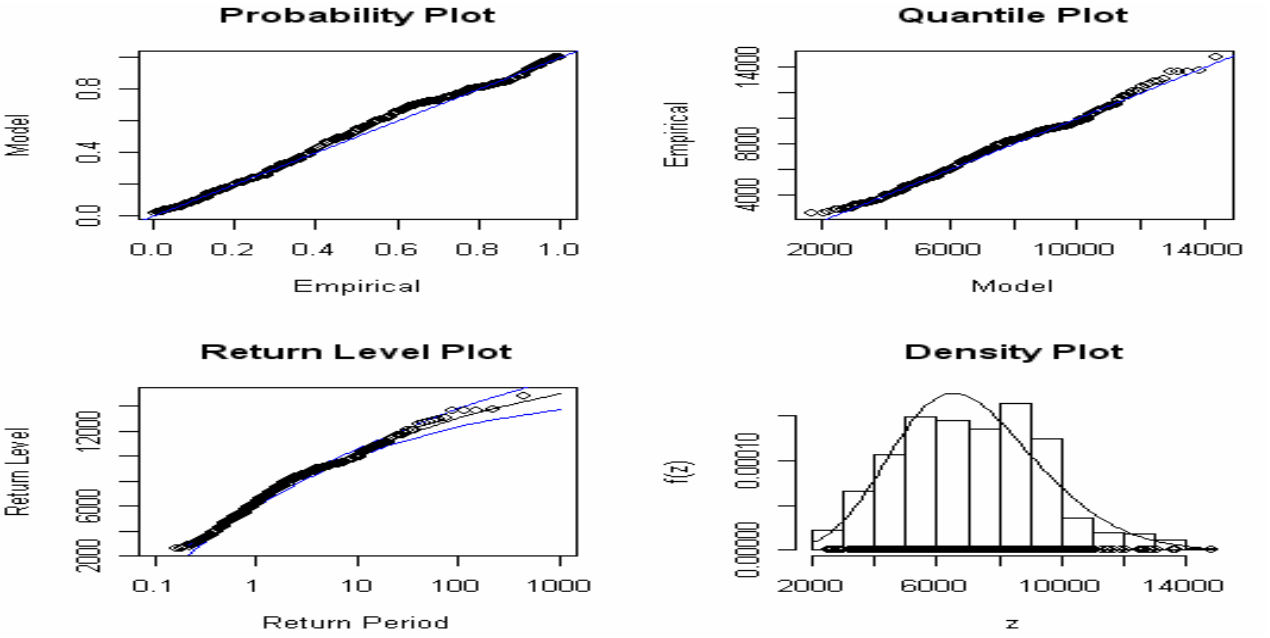


Fig.14. GEV fit diagnostics for Orsova monthly maximum discharge dataset (1958-2001). Quantile and return level plots are in mc/s.

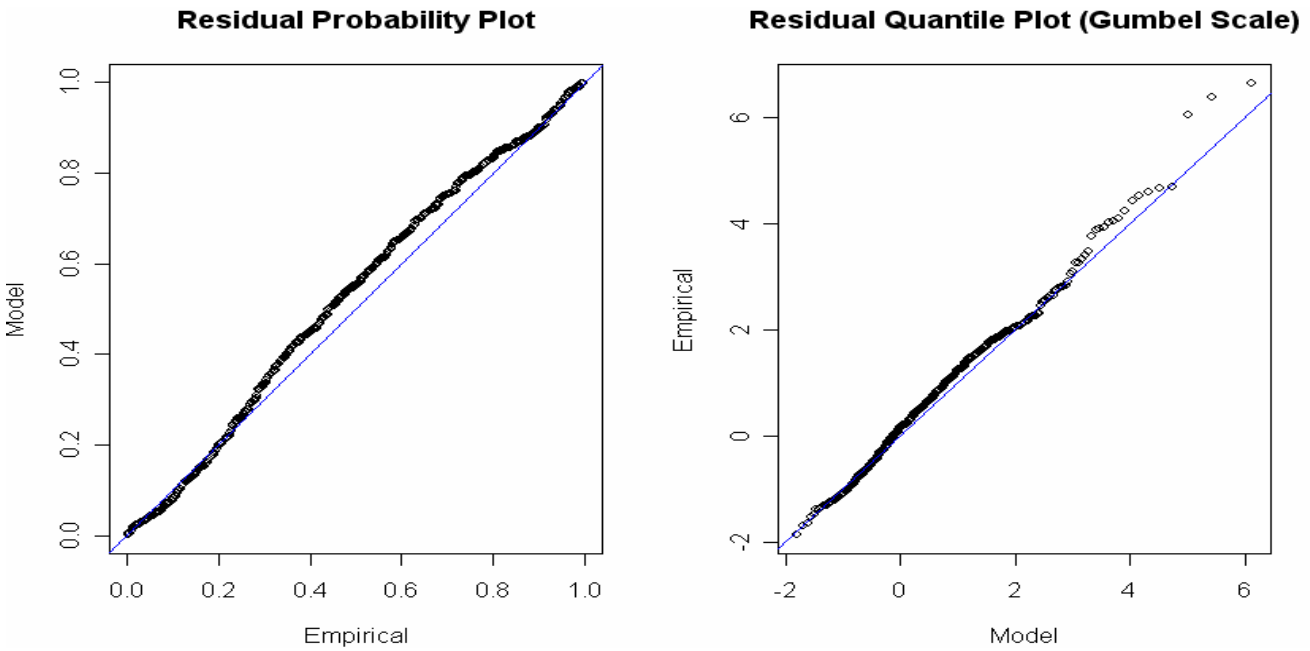


Fig. 15. GEV fit diagnostics for Orsova monthly maximum discharges dataset with PC1,…,,PC10 as covariates in the location parameter.

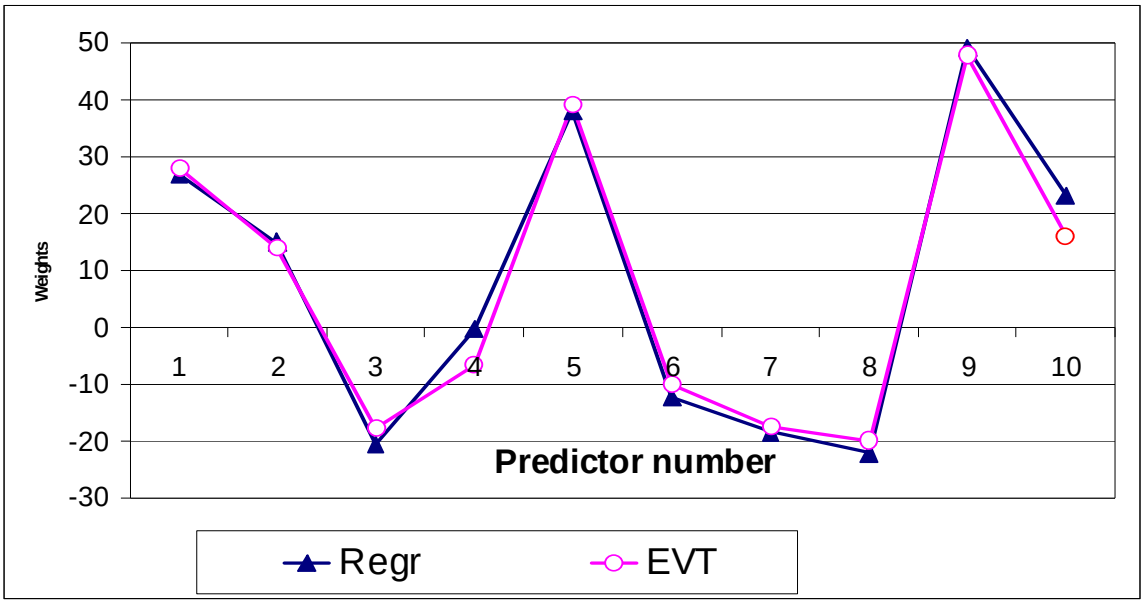


Fig. 16. Weights of PCs of MEOF decomposition of atmospheric circulation for a month before the predictand: monthly daily maximum of the discharge level of the Danube lower basin. CLR- Classical Linear Regression and EVT- extreme value theory.