ARCGIS SOFTWARE MODULE FOR CALCULATING THE S.P.I. VALUE

Zs. Magyari-Sáska[1]

ABSTRACT
Developing proper algorithm for different studies can have a major impact on research time, by decreasing the necessary manual work. If the algorithm is general on regional scale or with minor changes can be adapted for different locations its importance is even higher. The calculation of the S.P.I. value, characterizing the meteorological drought and excess of humidity, needs several steps and handles large amount of data. Because its calculus has a standard algorithm it's suitable to develop software to calculate it for every location of a raster layer. This study presents such a software module, capable to create raster layers with S.P.I. indices starting form locally measured precipitation, measure points' geographic location and altitude and digital elevation model.

*

1. INTRODUCTION

Drought or extreme moisture as natural risk phenomena were studied using several different approaches. Some of them use hydrological models (Haidu et al., 2004), others descriptive statistics regarding temperature and precipitation evolution (Croitoru et al., 2002). Towards the past period characterization, estimations for the future are also made in these studies. This is done using linear regression models or frequency analysis. While trend detection can estimate the value of the characteristic for a given temporal moment, frequency analysis determines the return period of an extreme (low or high) value (Haidu, 2003).

The S.P.I. (Standard Precipitation Index) is a well know measurement value for characterizing drought and moisture severity at different time scales (Moldovan et al., 2002). One of its advantages is that the only necessary data is the cumulated precipitation value for calculating the S.P.I. value. A comprehensive study (Lloyd-Hughes and Saunders, 2002) prove that the S.P.I. value can characterize the drought phenomena as well as the P.S.D.I. indices, which needs much more parameters and also some calibrated values.

The study of rainfall evolution, based on frequency analysis on different time scales represents one of the starting possibilities to determine extreme values characterizing drought or excess of humidity for given return periods. Such rainfall evolution study was made by Mercier and Haidu in 2003 for the Mediterranean region or Paul P. and David B. S. (2006) for the roumanian territory – Sulina Station.

Another possibility is to start the study from the generated S.P.I. maps of past periods, and the apply frequency analysis methodology on them. These kind of analysis was made for Thessaly region, Greece (Loukas and Vasiliades, 2004), were the calculated S.P.I. cell values were fitted on frequency distributions.

---

[1] *„Babeş-Bolyai" University, Faculty of Geography, 400006 Cluj-Napoca, Romania*

2. GENERAL CONCEPT

It's also well known that large amount of data can be handled in an effective way only if adequate algorithms are performed on them. Considering that the S.P.I. value has a spatial but also a temporal dimension, calculating tens of raster layers holding S.P.I. values for each point in manual way could be extremely time consuming.

The most important GIS software offers some automatic calculation methods (Modeler or Batch Processing in ArcGIS, or Model Maker in IDRISI), but they presents several important lacks:

- reduced possibility of parameterization (IDRISI)
- all parameters have to be entered in a manual way even if they are repeating or has a variable index in the file names (ArcGIS, Batch processing)
- they do not have algorithm controlling elements (loops, decision structures)

For developing specialized analysis software the use of script or programming languages is hardly advised. Scripts languages represent an advanced batch processing approach, but in generally they can't collaborate with other software environments and a user friendly application interface cannot be developed based on their commands. The collaborative environments are very important especially in research fields, because complex analysis cannot be done using just a single software product (Magyari, Haidu 2006). In such cases – using several softwares – the data transfer between applications can be done just by manual file transfers, which can be cumbersome. Analysis software that uses internal (inter-environmental) data transfers is much easier to handle and the researcher can focus on the studied issue in stead of counting the algorithm steps.

Such collaborative software can be developed just by using high level programming languages combined with the operating system capabilities.

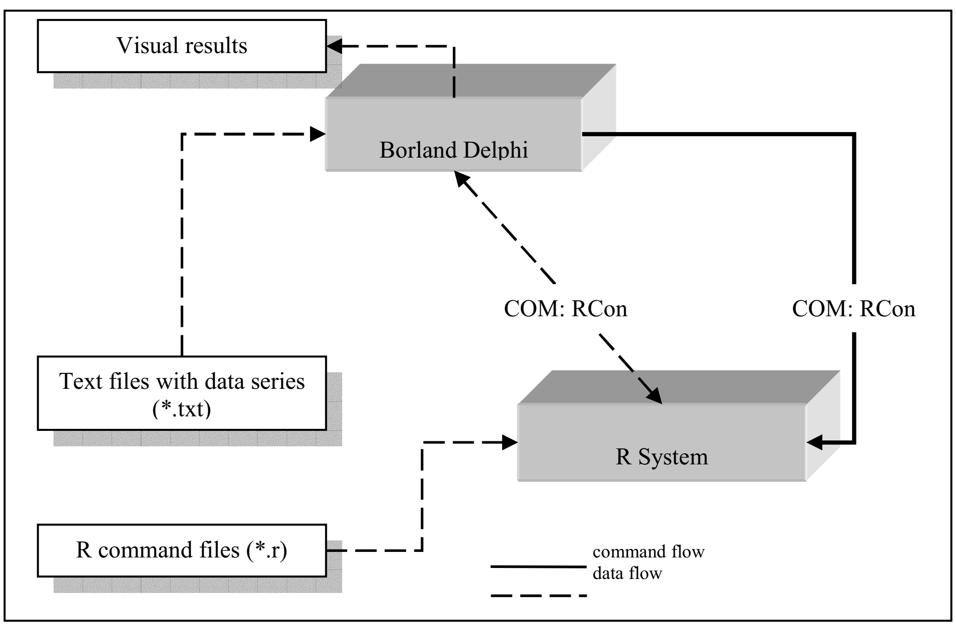

Fig 1. ***Conceptual design of the used collaborative environment***

This study presents an ArcGIS embedded methodology for calculating and estimating S.P.I. values by using probability distributions. Because such analysis use both spatial operations (existing in ArcGIS) and statistical calculus – which are not present in ArcGIS system –, a collaborative studying environment has to be developed. In this case the R System was chosen, based on two reasons: it's free being a GNU project and has an own programming language. This second argument is important for developing proper analysis function based on the existing ones. The R System's programming language with all the differences can be considered a C++ clone both in syntax and programming concept: all elements are considered to be objects.

The link between ArcGIS and R can be done using a COM (Common Object Model) component, RDCom freely available on Internet (www.sciviews.org) which makes possible also the MS Excel – R System command and data interchange. As mentioned before by using this component both data and command transfers can be done from ArcGIS towards the R System, and the analysis results can be accessible for ArcGIS environment.

3. THE APPLIED ALGORITHM

The aim of the study was to develop and implement an algorithm which easily can calculate and estimate future S.P.I. values for a chosen time period based on DEM and locally measured precipitation values.

The first level steps of the algorithm are:

- determining the precipitation field for the study area
- calculating the S.P.I. values for each chosen time segment
- determining the S.P.I. values for different return periods

Step 1. Determining the precipitation field

Different interpolation methods can be used for specializing locally measured values. In this study the multiple regression method was used, as the application should be used on poorly gauged regions, with different topographic properties, and prior investigation and other interpolation methods presented unsatisfactory results.

```
intp<-function(prec)
{
 expo=scan(file="expo.txt")
 pozx=scan(file="pozx.txt")
 pozy=scan(file="pozy.txt")
 inalt=scan(file="dem.txt")
 d=data.frame(expo,pozx,pozy,inalt,prec)

 teo=lm(prec~pozx+pozy+inalt+expo+I(pozx^2)+I(pozy^2)+I(inalt^2)+I(expo^2)+I(pozx*pozy)
 +I(pozx*inalt)+I(pozy*inalt))

 model=stepAIC(teo)

ln=c("(Intercept)","pozx","pozy","inalt","expo","I(pozx^2)","I(pozy^2)","I(inalt^2)","I(expo^2)",
"I(pozx * pozy)","I(pozx * inalt)","I(pozy * inalt)")

va=rep(0,12)
 for (i in 1:length(model$coefficients))
 {
  for (j in 1:12)
   {
    if (names(model$coefficients[i])==ln[j])
     {
      va[j]=model$coefficients[i]
     }
   }
  }
```

Fig 2. ***Determining the most suitable multiple linear regression model (R System)***

In the interpolation algorithm 4 major factors was considered: altitude, geographic position and aspect. Due to the fact that the first three values are in continuous space while the last one represents the North aspect using two separate interval (0-22.5 and 337.5-360, if we consider 8 geographic directions), for a multiple linear regression is necessary to translate the whole aspect range in continuous, successive intervals. The transformation used shifts the aspect values from 0 to 337.5 by 22.5 towards 360, and inserts the values between 337.5 and 360 at the beginning of the resulted interval.

The multiple linear regression equation is optimized by R System stepAIC command contained in the MASS package. For optimization the regression model should be transmitted as a parameter. The optimization process eliminates the model terms one by one. By this the algorithm determines the minimum AIC value, showing the resulted, final model. As the command result the correlation values between the tags are also displayed. The lower values indicate a less correlation between the tags, which is preferable because it means that there are less information overlap between the tags considering their importance on the model.

Starting from the optimized model the calculation of precipitation values based on the mentioned factors is done in R System using the exported ASCII raster files. The situation of calculated negative precipitation values also should be considered. The most simple way is to assign a zero value in this cases could not represent the differences between this regions, so a new approach was used. All those cells having negative precipitation values got new values which is the reverse on their modulus.

$$p = \begin{cases} p, \text{ for } p \geq 0 \\ \frac{1}{|p|}, \text{ for } p < 0 \end{cases}$$

Using the described method several raster layers are calculated, each one for a specified different time period (in general for 3, 6, 9, 12 or 24 months).

Step 2. Calculating the S.P.I. values

The S.P.I. value calculation is based on Gamma distribution. First the distribution is fitted on the obtained, corresponding cell values. In this case the distribution fitting was done using method of moment's methodology (Ricci, 2005). This method was chosen because the fitdist command in R System – which uses maximum likelihood method for distribution fitting – in some cases produces error (especially when some values are close to zero an the other ones has considerable higher values). After that the gamma probability values are calculated for every considered value. To obtain a normalized indices value the formulae use is developed by Abramowitz and Stegun (1964)

$$SPI = -(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3})$$ when the calculated probability is in (0,0.5] interval

and

$$SPI = +(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3})$$ when the calculated probability is in (0.5-1) interval

where

$$t = \sqrt{\ln[\frac{1}{H(x)^2}]}$$, for probabilities in (0,0.5] interval

and

$$t = \sqrt{\ln[\frac{1}{1 - H(x)^2}]}$$, for probabilities in (0.5,1) interval

The c and d parameters has the following values:

c_0=2.515517, c_1=0.802853, c_2=0.010328, d_1=1.432788. d_2=0.189269, d_3=0.001308

For each selected time period the precipitation values for corresponding cells are cumulated and the S.P.I. values are calculated. In such manner several ASCII raster files are created holding the S.P.I. values.

Step 3. Calculating S.P.I. values for different return periods

Considering Gamma distribution the most suitable to characterize the precipitation at regional scale, this distribution was used to determine S.P.I. values for different return periods. This operation is done also in R System, using the files created in the earlier step. The resulted file is completed with a proper header, taken form the digital elevation model ASCII file, and presented in ArcGIS.

4. THE RESULTED SOFTWARE MODULE

The described algorithm was implemented in Visual Basic for Applications in ArcGIS. To use it you should hold an ArcGIS project file containing this macro. Before using it the DEM raster of the study region should be present in ArcMAP.

For using it, first the working directory should be specified. In this step the existence of a few text files are verified. These files should contain the measure point's altitude, transformed aspect and geographic position. The lack of it will be announced by their names showed in red, otherwise the names appears in green. From the specified directory all the files containing monthly cumulated precipitation values are collected and presented in a list box. These files are selected based on their name format: all of them should have an X as the first letter in their names. There should be at least 12 precipitation files, because the linear regression model uses 12 factors.

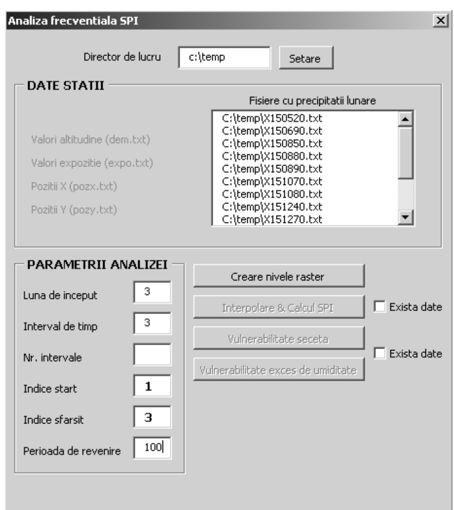

Fig 3. ***User interface of the developed ArcGIS module***

If everything is found the module enables the first analysis button, which creates the necessary raster layers (*Creare nivele raster/Creating raster layers in English*). These layers are:

- the aspect layer
- the transformed aspect layer
- a layer containing for every cell it's center X position
- a layer containing for every cell it's center Y position

The last three layers together with the DEM are the automatically exported in ASCII raster format. Because these layers are common for every study regarding a specified region their re-creation can be omitted by checking the first check button (*Exista date/Existing data in English*).

```
ListBox1.Clear

Dim FSO As Object
Set FSO = CreateObject("Scripting.FileSystemObject")
Dim Fl As Object

strfile = Dir(cale & "\X*.txt", vbNormal)
Do While strfile <> ""
If strfile <> "." And strfile <> ".." Then
Set Fl = FSO.GetFile(cale & "\" & strfile)
ListBox1.AddItem (Fl)
Set Fl = Nothing
End If
strfile = Dir()
Loop
```

Fig 4. ***Populating the list box containing monthly cumulated precipitation values (VBA in ArcMap)***

Then, with a new button (*Interpolare si calcul S.P.I./Interpolation and S.P.I. calculus in English*) the first two steps presented in previous chapter are realized: precipitation interpolation and S.P.I. value calculus. These analyses are done based on two parameters defining the study period. The first values specifies the starting month (*Luna de inceput/Starting month in English*), while the second one the length of a study period (*Interval de timp/Time period in English*). If these calculus are done they can be used for determining two hazard types, and their existence can be notified for the module by checking the second check button in the window.

```
cmd = "f=1:" & Str(ListBox1.ListCount + 1)
RLink.EvaluateNoReturn (cmd)

For i = 0 To ListBox1.ListCount - 1
 cmd = "f[" & Str(i + 1) & "]='" & ListBox1.List(i) & "'"
 cmd = Replace(cmd, "\", "\\")
 RLink.EvaluateNoReturn (cmd)
Next

Dim st As String
Dim pe As String

st = TextBox2.Text
pe = TextBox3.Text

cmd = "h=recalc(" & Str(ListBox1.ListCount) & "," & st & "," & pe & ")"
RLink.EvaluateNoReturn (cmd)
```

Fig 5. ***Calling R System commads form VBA environment in ArcMAP***

The final steps can determine the drought hazard (*Vulnerabilitate seceta/Drought hazard in English*) or excess of humidity hazard (*Vulnerabilitate exes de umiditate/Extreme humidity hazard in English*). For creating these raster layers the desired return period (*Perioada de revenire/Return period in English*) and two indices should be specified which represents the first (*Indice start/First indices in English*) and the last (*Indice sfarsit/Last indices in English*) S.P.I. files indices. The maximum possible value can be determined from *Nr. intervale/Inreval number in English* fields.

The resulted a raster layer is reclassified according to the standard S.P.I. intervals and presented in ArcMAP.

ACKNOWLEDGEMENT. The study was financed from CNCSIS Research Project No. 492/01.10.2007.

REFERENCES

Abramowitz, M. and Stegun, I. A., (1964, 1972), Handbook of Mathematical Functions, Dover Publications, New York

Croitoru A.E., Hauer E., Mihăilescu M., (2002), Riscuri determinate de temperaturile extreme şi de cantităţile maxime de precipitaţii căzute în 24 de ore în nord-vestul României, Riscuri şi catastrofe, Casa Cărţii de Ştiinţă, Cluj-Napoca

Haidu I. (2002), Analiză de frecvenţă şi evaluare cantitativă a riscurilor, Riscuri şi catastrofe, Casa Cărţii de Ştiinţă, Cluj-Napoca, 2002

Haidu I., Sorocovschi V., Imecs Z. (2003), Utilizarea S.I.G. pentru estimarea riscului de producere a evenimentelor extreme: Excesul de umiditate şi secetă din Câmpia Transilvaniei, Riscuri şi catastrofe, Casa Cărţii de Ştiinţă, Cluj-Napoca

Loukas A., Vasiliades L. (2004), Probabilistic analysis of drought spatiotemporal characteristics in Thessaly region, Greece, Natural Hazards and Earth System Sciences

Lloyd-Hughes B., Saunders M.A. (2002), A drought climatology for Europe, International Journal of Climataology, vol. 22

Magyari-Sáska Zs., Haidu I. (2006), Posibilităţi de modelare spaţială în mediu programat, Analele ştiinţifice ale Universităţii „Al. I. Cuza" din Iaşi, Geografie, Tomul LII, 2006, ISSN 1223 5334, 9 pagini (57-65)

Mercier J.-L., Haidu I. (2003) Reconsidering the mediteranean rainfall: the case of the ecarts between the monthly extremes and monthly totals. Vol. Hydrology of the Mediterranean and Semiarid Regions, Montppellier, IAHS Press, 6 p. CD.

Moldovan F., Sorocovschi V., Holobâcă I. (2002), Deficitul pluviometric ca fenomen climatic de risc în Depresiunea Transilvaniei, Riscuri şi catastrofe, Casa Cărţii de Ştiinţă, Cluj-Napoca

Paul P., David B. S., (2007), Analysis of Historical Precipitaion 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

Ricci V. (2005), Fitting distributions with R, cran.r-project.org/doc/contrib/Ricci-distributions-en.pdf

***, S.P.I. Defined, http://ccc.atmos.colostate.edu/pub/SPI.pdf

***, Programming in ArcGIS using Visual Basic for Application, web.mit.edu/gis/www/programming_iap2006