Processing of data for seismic microzoning

SMLAC (UNESCO/IGCP project 487, ICTP Network NET-58)

SMZ

Third version, September 2011

L.Alvarez, CENAIS

Previous versions

First version, December 2008

Second version, July 2009

Index

Introduction

During a run of jobs job2D, job2D64, job2Db and job2D64b it is created a compressed (profile.dat_P4.zip) file with a lot of information about frequency dependent ground motion (GM) parameters (see annex 1) and another file (profile.correl.zip) with information about 2D GM parameters, 2D/1D GM parameters ratio and a non parametric correlation analysis of the last ones (see annex 2). The frequency dependent information can be used for for preparing curves for classification, the 2D and 2D/1D information are suitable of space plotting [all of the files are of the kind (x,y,z)], while the correlation analysis is useful to pick up the variables that are really independent from the GM parameters. In present document are discussed all these possibilities. Plotting of results is done mainly online with GMT or gnuplot, but it is given the possibility of preparing some plots using a demo version of IDL.
Warning: In the following, the sample input files are written with "Courier" fonts. When a line is so long for page it is continued using French indention.

a) Classification of frequency dependent ground motion parameters

Here are explained how to prepare the curves to classify with Windows programs P4, the main features of this program and how to process the results of the classification with seismic microzoning purposes.

a.1) Programs that do the work

Below will be described the programs that extract the data from initial format and prepares it to process with P4, as well as those that process the results of a P4 classification.

"cu2ma.f" is the program for preparing input matrices for P4 program

It reads a set of files with GM data corresponding to sites in a profile and the data for calculating the coordinates of its points. Then, it prepares the initial matrices for PROGNOSIS (version P4 for Windows) for making a logical-combinatorial classification. During runtime it creates 2 subdirectories (“ma” and “ps”) where it places the results.
The program cans control sampling of data, both in sites and in frequency, and smooth curves with a moving average window. Additionally given the data about profile and sites' separation it fixes the coordinates of each site.
The initial data is of 2 kinds:

1) File cu2ma.inp that you have to prepare with data about your profiles and classification options
2) The files contained in the "name_of_profile.dat_P4.zip" file generated in a run of "job2D". They are files with eir, rsr and other data listed in the "fdplota.*" files (included in the same compressed file) that have to be converted to ".ma" files. They have the name “profile.code”. The "fdplota" names are given in a subroutine of the proper program
3) A file “P4inp.txt” with a listing of codes corresponding to the different input files.

Structure of the "cu2ma.inp" file:

```
Use   of   coordinates(0-no,1-yes),sampling   interval(site,frequency),log   of
      values(0-no,1-yes), smoothing average window length (points), value for
      adjusting the theresholds εt (formula A-5,Annex 3)
1 2 5 0 5 5.
Data for coordinates (x0,y0,az,delta)
522289  188198 55.6  84
```

Second line: parameters governing process (5 integers in free format)

icoor (0 - no coordinates of points, 1 - with coordinates
isamp - interval of sampling of initial data in sites' position
isampf - interval of sampling of initial data in frequency
ilogdata - use of logarithm of amplitude data (0 - no, 1 - yes)
ismooth - length of moving average window for smoothing (odd value between 1 and 9, 1 means no smoothing)

Fourth line: Data for coordinates of sites (4 numbers in free format)

x0 - plane X coordinate of profile beginning
y0 - plane Y coordinate of profile beginning
az - azimuth (from North) of profile trace
delta - distance between sites in meters calculated by pfdg9m (look for this name in lines 10-25 of pfdg9.pri). -warning: original distance is in km.

Output:

1) "name_of_file.ma" files that contain the initial matrices for PROGNOSIS, where name_of_file is the name of the corresponding input file with data (that is included in "name_of_profile.dat_P4.zip"), in the subdirectory “ma”. The matrices has as variables the values of frequency or period, and as objects the values of the field (GM parameter).
2) "name_of_file.ps" files with a plot of the initial and smoothed and curves in the subdirectory “ps”. Plots are done with “gnuplot”
3) "name_of_file.estad" files in subdirectory “est”. They contain statistics for each variable (extremes, average, median, absolute first moment, standard deviation, variance, skewness and kurtosis)
4) file ”P4inp.tx” with list of GM codes corresponding to the different generated matrices

The process is profile-dependent, and at this stage you can obtain only individual input matrices for classification by profile. If you want to classify the data for all the profile, once you finish the calculation of the matrices for all the profiles you have to join similar data matrices by using the program "join_ma". In this case is mandatory to use the same value of sampling in frequency for all the profiles. The value of therresholds ε_t for each variable are calculated as (max-min)/last_value_in_2nd_line.

Warning: In this process, the presence of dots ('.') in names of files is conserved. But take into account that you can not use this names for classifying the curves with P4, because it is a Windows program and does not accepts dots in names except the one that separate name from extension.

"join_ma" is the program for joining several initial matrices ".ma" for P4

It reads the input matrices for classification and join them in only one. It requires that all the matrices are formed for the same range of values of frequency (x coordinates). If no, it stops and prints a warning message.

Input data:

1) File 'join_ma.inp' with list of files profiles which ".ma" files should be joined

```
5. value for adjusting the theresholds εt (formula A-5, Annex 3)
list of profiles to process
B11
B12
B13
B14
```

2) File 'P4inp.txt' with the list of codes characterizing the different matrices (the same for all the profiles). It is generated automatically by "cu2ma" program

```
f0.zac.eir_lg
tra2d.rsp
rad2d.rsp
ver2d.rsp
```

3) Files '.ma' to be joined for all the profiles

Output:

1. Files 'JOIN_code.ma' with all the information from initial files for the same "code". This file has no intermediate dots ('.') into the name and can be used directly to classify with Windows program P4.
2. Files "JOIN_code.estad" files in subdirectory "est". They contain statistics for each variable (extremes, average, median, absolute first moment, standard deviation, variance, skewness and kurtosis)

Program makes an alert when the total number of sites is greater than 350, because present version of program P4 (number 10) fails when the number of sites is about 400 due to limitations of the graphical representation that it uses. If that is the case, you have to reduce the total number of sites, by sampling, removing of profiles in the run, or manually deleting sites from the resulting '.ma' file. In new versions of program P4, this problem should be solved.

Classification with program P4

The program P4 in its present version (P4r10.exe) is able to classify until 400 curves. It uses as input the files 'name.ma' generated by programs 'cu2ma' and 'join_ma'. I does not accept dots in 'name'. Normally, outputs of cu2ma should contains dots in 'name'; then, if you want to classify these outputs you have to rename them, use "check_ma" (see below) for do it. Program 'join_ma' gives outputs without these intermediate dots. The output of a classification program is in 'name.htx'.

Program P4 of course run in Windows, but you can run it in Linux using

```
wine "path"/P4r10.exe
```

("wine" is present in all Linux distributions). If you run P4 in Linux it is tested that accepts input files in UNIX format, but the outputs are in DOS format. Nevertheless these output can be used as inputs for the other programs of this package without format conversion. The use of programs that creates virtual machines on Linux and Mac could be used to run this program in a Windows environment. The usage of program P4 is explained in annex 3.

"check_ma" corrects the name of input matrices to accepted ones by P4

Program "cu2ma" follows the naming convention of all the programs of the package of synthetic seismograms' calculation. Nevertheless, P4 is a Window's

program that only accept one dot “.” in the input files, i.e. “.ma”. Then, you need to rename any file that you want to process by P4 if it contains more 'dots in name. It is made by program check_ma. It lists all the files “.ma” present in current directory and change the names of those containing multiple dots by changing “.” by “_”, letting only the one of “.ma”. If the original name is correct, programs do not alter it. A report of what has been done appears in check.pri.

"plotclas" is the program for preparing plots of the results of the classification of curves with program P4

Once you have run P4 and saved your results in an '.htx' file, you can run this program to pass to a graphical representation of results. It uses as inputs the input and output of a run of P4 program ('.ma' and '.htx'). You has to give interactively what is called "name of the problem" (='prob'), which is the part of the name of P4 files before the '.ma', and the program searches for the corresponding data files

Inputs: Through file "plotclas.inp" or interactive. Both provide:

```
prueba
bayamo.xy
5
```

1) Name of the problem given interactively: "prob"
2) Name of a contour file for GMT plotting of results
3) Diameter of sites symbols in the plot'

program searches for additional files:

4) Initial matrix for Prognosis : "prob.ma"
5) Results of classification : "prob.HTX"

The outputs have fixed names, formed from 'prob', and are used as inputs of 2 IDL programs: ('plmap1' and 'pclamu')

Outputs:

1) prob_post.dat - coordinates of site and number of group in which it was classified
2) prob_punt.dat - coordinates of site and its ordinal number in initial matrix
3) probMpost.dat - modification of file prob_post after the on-line subdivision of some groups, which constitutes a revised classification by other criteria.

This files are used to prepare some maps of input and output distribution with idl program 'plmap1'

4) probC.idl - average curve for each group; first block - number of groups and frequencies,integer codes; second block - frequencies; third block - number of group, ordinal number, coordinates, data of curves
5) prob(1:5)Gnnn.idl - where nnn is the number of the group, one file per group. First block - number of sites in the group, number of frequencies, number of site to which group corresponds, number of the group second block – frequencies, third block - number, name and coordinates of the site, curve
6) probM.idl - modification of file probC.idl after the on-line subdivision of some groups, which constitutes a revised classification by other criteria

These files are for plotted with idl program 'pclamu'

7) pcmu.inp - input file for plotting with IDL program 'pclamu'
8) pclas.pri - information about compilation
9) fdtriangul.par - input for online plotting of a map with groups distribution

using perl script "fdtriangul.pl"

10) prob.P4.ps - a plot of group number classification on a map

Input file:

The program have the option of groups' subdivision, which may be useful when you classify using the connected components algorithm of P4. In this case you have to edit the file 'pcmu.inp' for adjusting the total number of graphics to plot (line 4). If instead of groups' subdivision you need groups' joining, you have to use program 'joingroup' .

'plotclas1' is a modification of 'plotclas' for batch processing the results of P4 grouping

Once you have processed all the matrices with P4 and you have saved the results of this process in “htx” format (“.htx” if processed in Linux and “.HTX” if processed in Windows) you can run plotclass1 for obtaining related plots. If you have used program microMapP4 (see below), the matrices and the results of classification should be in the directory “P4/matP4”. You will need a contour file for GMT plotting of results. The input file (plotclas1.inp) is:

```
base map for plotting
bayamo.xy
5       dimension of posts in plot (always integer)
input matrices
prueba_D.ma
prueba_R.ma
```

The program creates a directory for processing the results of classification (the .htx files) for each matrix (no matter if they have extension .htx or .HTX). In each directory it will put all the data for making auxiliary plots; the ones explained for “plotclas” program plus two for making a one page plot of average curves for all groups (“gnp.job” and “gnp.data”). In the directory P4/matP4 it will be created a script “pclas.plot” that should be running by “sh plotclas.plot”. This script makes automatically the plots with GMT through perl program “fdtriangul” (see below) and “gnuplot” and start interactive sessions of IDL plotting. When IDL is started each time, you have to write “pclamu” for plotting and exit” for returning control to Linux environment.

Warning: For using IDL program in a batch file it is necessary to change its settings

- edit the “idl_setup” file (looks for it in your environment file) and “comment” the line with the alias to “idl” program
- add a line in your environment file with the path to “idl” program

It creates also the input for htx_contour program that marks a boundary for groups in map (see bellow) and runs it.

'joingroup' is used for joining groups

In process of classification you may obtain a bigger quantity of groups as you desire to process on, or perhaps you obtain a lot of groups with little elements. As microzoning is a process of generalization, you may be interested in joining groups that in the process of classification resulted separated. The criteria for joining is at individual consideration, and may be due to geophysical, geological or geographical considerations and so on.

It uses an input file 'join.inp' that contains:

```
bayamo.xy                                   :contour file for GMT plotting
6                                :dimension of numbers in GMT plotting (integer)
zonacionMpost.dat                           :input post file
2                                :quantity of groups to join
1 2                              :corresponding ordinal number of groups
2                                :continue in the same way
3 4
--------------
6
5 6 7 8 9 10
```

The outputs are:

pcmu.inp - input file for plotting curves with idl program 'pclamu'. You have to edit this file for placing the real number of files to plot (line 4)

probG###.idl - files with curves corresponding to new groups formed, when ### is a consecutive number after the last of the original groups numeration made by 'plotclas'

probMJpost.dat - modification of initial 'probMpost.dat' file after joining groups

probCJ.idl - modification of 'probC.idl' (of the 'plotclass' run) file after joining groups

fdtriangul.par - input for online plotting of a map with groups distribution using perl script "fdtriangul.pl"

probMJpost.dat.ps - a plot of the numbers of groups in the final version

'joingroup1' allows online plotting of IDL graphs

It uses the same input file that “joingroup”. You need to made the changes in IDL settings reflected for program “plotclas1” to allow the online use of IDL. The program crates a new sub-directory “join” where place its results. It also runs online the program “htx_contour” (see bellow) that marks a boundary for groups in a map.

a.2) Graphical IDL programs

These programs are used for plotting results of the classification in an independent run. The IDL should be called by command “idl” and then the control is passed to this plotting system. Each program is called independently. In the standard IDL installation it is not possible to run these programs in other way, and then, they can't be included in a script. Nevertheless you can modify the IDL installation and then these programs can be run from a script. For do it you need:

1) to comment the alias to idl present in file "bin/idl_setup" of install directory

2) to make a link to idl program in "/usr/bin", e.g.:

`sudo ln -s idl_install_directory/bin/idl /usr/bin/idl`

Then, your "idl" main program is considered a system one and runs normally from scripts. But each time the control is passed to idl you need to type in the idl prompt the name of the idl program to be executed (e.g.: pclamu) and after finishing the program to type "exit" for returning to main prompt and continue running the script. All programs in version 3 are adapted to this way of running, and before passing the control to idl program it is written a message of the kind:

!!!WARNING!!!!! -> TYPE pclamu AND exit AFTER THE idl PROMPT

Until present it cant be solved the problem of passing these names to idl program directly from the script.

pclamu

Plotting of multiple graphics with results of classification in a single sheet. It processes the files '.idl' included in the list "pcmu.inp"' that are the output of programs 'plotclas' and 'joingroup'.

Structure of pcmu.inp

```
 0                    orientation (0 - portrait, 1=landscape)
 2                     number of columns
 3                     number of rows
 6                     number of graphics to plot
qqq                    name of the output file   (less than 10 characters)
below put the names of the graphics and selection of axes titles
qqqMJ.idl
1 0   xtit(0-none,1-Freq,2-Per),ytit(0-none,1-rsr,2-eir,3-lg(EIR),4-Vg)
.......
```

With the 3 first parameters you can configure your output graphic, that will have the name 'qqq.ps'. You can experience different configurations changing the name of output file. The numbers of graphics to plot is only correct in the case that you run plotclas without subdivision, but has to be corrected if you use subdivision or in a run of 'joingroup'.

plmap1

This program plots the geographic contour of study region plus some posting information (x,y,number) for points into the region. The input files have the same format as corresponding files for making the same graphic with Windows program SURFER. The default input file is 'plmap1.inp'

The file 'plmap1.inp' contains 9 lines:

1) Minimum Longitude, Maximum Longitude, Minimum Latitude, Maximum Latitude. These are the coordinate limits for the map (free format)
2) Name of the file that contains geographical contours to be plotted (if no contours are desired put 'none'). This file has a format slightly modified from the '.bln' used by SURFER program for Windows. Each contour in the files appears as
   - Number of points(N), fill_option(ifill: 0-no,1-yes)
   - N lines with coordinates of points (longitude,latitude)
3) Name of the file with posting information. This file has the format of the '.dat' files created by program 'plotclas' or the corresponding '.dat' of program SURFER; i.e., an unlimited quantity of lines with (longitude, latitude, post_information) where post_information is an integer. If no post are desired put 'none'.
4) Name of the file that contains contours of profiles to be plotted. This file has a format slightly modified from the '.bln' for SURFER as indicated in point (2). If no profiles are desired put 'none'.
5) A title for the graphic. You can supply the title or let it blank by putting an empty line; if you put the character chain 'auto', the program creates one with the name of the posting file; if you put the character chain 'none', the program puts no title.
6) Interval of ticks in graphic, measured in data units

7) The size of the post over map. The normal is between 0.6 and 0.8. If you put 0, the program uses 0.6
8) The position of post with respect to X coordinate (0 means left justified, 0.5 means centred and 1 means right justified). The centred option means to centrer the bottom centre point of the post.
9) The displacement in Y coordinates of the post, positive down (same units as data). Use this parameter for centring posts in Y. Run the program with 0, measure over the screen the displacement of longitude that you need, and rerun the program with this value

Example:

```
-78.5 -75.0 19.00 20.3                              :limits of region
cuba2.bln
per.post
per.bln
Ewing'60
0.5              :interval of ticks in map (data units)
1.0               :size of posts
1.                :post - X position (0 - left ,0.5 - centre ,1 right)
0.                :post - Y displacement (positive down), data units
```

During running the program determines the real maximum and minimum coordinates of data. You can use this information to vary the parameters of the first line in order to obtain a more adequate graphic representation.

Warning: Programs 'plotclas' and 'joingroup' do not create the input file 'plmap1.inp'. You have to create it giving the geographical, profiles and other information that you want. If you wand to add another post information, the only way in present version is to add it at the end of the post file created by 'plotclas' or 'joingroup' programs

a.3) General procedure for classification:

- Run 'cu2ma' for creating input files for P4 program from the data calculated in a job2D run, contained in 'profile.dat_P4.zip' file, in order to add coordinates of the sites, to sample and smooth in frequency and, if necessary, to sample sites.
- Once it has been done for all the profiles, take the decision about which curves you will use for classification with microzoning purposes
- Join all the initial matrices for one kind of curve using 'join_ma'
- Run 'plotclas' without groups' subdivision and plot the results using idl programs 'plmap1' and 'pclamu'
- Analyse the results and take the decision of groups subdivision or joining. It is a good help to analyse together the plots of curves for groups and the map of sites' distribution.
- For subdivision rerun 'plotclass' and for joining run 'joingroup'. Follow this order because 'plotclas' uses the initial data ('.htx') from classification, while 'joingroup' uses as input the results of 'plotclas'
- Plot the results using idl programs 'plmap1' and 'pclamu'

b) 2D plotting of GM parameters

The following procedure starts from the point that you have calculated synthetic seismograms with an adequate space distribution of sites that allows to use space interpolation techniques. If that is not your case, the maps that

you will obtain with plotted fields will not be reliable.

b.1) Programs that do the work

In file “profile.correl.zip” the GM data is presented in several files in the form of tables as function of site number along profile. Below will be described the programs that extract the data from initial format and prepares it to process with space interpolation.

“mapesce” is used for generating (x,y,z) data for plotting from the tables

```
Name of 2D scenario file
B11.esce
Data for coordinates (x0,y0,az)
522289  188198 55.6Input: “mapesce.par” with four lines
```

x0 – initial “x” coordinate
y0 – initial “y” coordinate
az – azimuth (from North)

Output:

- one (x,y,z) file for each column of the tables processed:
  48 files with 2D GM data, placed in directory “2D”
  54 files with 2D/1D ratio of GM data, placed in directory “GMrat”
  24 files with 2D/1D ratios of H/V relations, placed in directory “HVrat”
  The output files have the name “profile.code.dat”, where “code” is formed in relation to the data contents.
- a file with the list of codes used in the different cases, placed in the corresponding directory (“GM.txt”, “GMrat.txt”,”HVrat.txt”)

“joinmap” is used for joining (x,y,z) files

Place the “.dat” files created by directory for all profiles in one directory, and add the common “.txt” file. It means that you have to form the three directories (“2D”, “GMrat”,“HVrat”) each of one containing the “.dat” files obtained for all profiles (“profile.code.dat”). The program will create a set of files “all.code.dat” according to the data included in the corresponding file (“GM.txt”, “GMrat.txt”,”HVrat.txt”). It has as input data the file joinmap.par, with the following structure:

```
Parameters for gridding and maps' preparation
0.05       diameter of sites symbol in plot
0.8        tension in minimum curvature
1200       Search distance for nearest neighbour (m)
3          number of sectors for nearest neighbour
400        Distance for Delaunay triangulation (m)
geographic contours file
bayamo.xy
list of profiles names to process
B11
......
......
B33
```

The first part are the input for defining maps preparation with the “fdtriangul.pl” perl script (see above). The program creates an input for this

script (“fdtriangul.par.code”) for each one of the joined “all.code.dat”. The list of profiles should finish without blank lines below (if it is not done with this care, program will stop).

“fdtriangul.pl” is a perl script for plotting (x,y,z) files with GMT

The GMT programming tool allow to process data for making different kinds of plots, including maps with a multiplicity of geographical projections. In the case of seismic microzoning, due to scale it is convenient to use planar coordinates, then the basic unit is the meter (m) and if you want to plot geographical accidents you should provide them in a file of the kind (x,y) - (bayamo.xy in the joinmap.par example). GMT includes several tens of programs that have to be called in command line, while the parameters that they used should be passed as arguments in this command line. It is facilitated by the use of perl programming. Between these programs there are several that allows to create a regular grid from unequally spaced observation data points. The script “fdtriangul.pl” creates 4 maps and 3 data files for each input “all.code.dat” or a map for each “JOIN_code.dat”:

code.XY.ps – a plot of the study area with sites where GM parameters have been calculated

code.DT.ps – a plot of the input data in a regular grid obtained by a Delaunay triangulation procedure

code.DT.xyz – an ASCII file with the result of gridding with the Delaunay triangulation procedure. Points with no data are filled with the value 1.10141e+38 as it is done by SURFER for Windows.

code.MC.xyz – an ASCII file with the result of gridding with a plot of the input data in a regular grid obtained with the minimum curvature interpolation procedure

code.MC.ps – the minimum curvature interpolation procedure

code.NN.ps – a plot of the input data in a regular grid obtained by a nearest neighbour interpolation procedure

code.NN.xyz – an ASCII file with the result of gridding with the nearest neighbour interpolation procedure. Points with no data are filled with the value 1.10141e+38 as it is done by SURFER for Windows.

code.estad – an ASCII file with main statistical characteristics of Z component (GM parameter) of input file

In file “joinmap.par” you should give the basic parameters for interpolation. Try several values using only one “all.code.dat” file, and then extent your selection to all the desired files.

JOIN_code.dat.P4.ps – a plot of the study area with sites used in classification

In file “join.inp” you should give the name of a contour file to use in this plot and the dimension of the numbers of groups to be placed in the map.

The file with geographical accidents contain:

```
> any comment, followed by first contour
x1,y1
-------
xN,yN
> any comment, followed by second contour
x1,y1
-------
```

```
xM,yM
etc.
```

You can put as much contours as you want, but they should be separated always by a line that begins with “>”.

For running “fdtriangul” script you should issue the command

`fdtriangul.pl fdtriangul.par.code`

The input file is:

```
input parameters for plotting amplifications with the model
-----------------------------------------------------------
0          site symbols (1 – print,0 - not print)
5          dimension of group number in P4 plotting
0          (0=gridding, 1=plot P4 grouping)
0.05       diameter of sites symbol
0.8        tension in minimum curvature
1200       Search distance for nearest neighbour (m)
3          number of sectors for nearest neighbour
400        Distance for Delaunay triangulation (m)
esc.Dmax2_rad                                    PostScript base filename
all2Desc.Dmax2_rad.dat                           Data file
bayamo.xy                                        Contour file
```

It is used both for scenario and P4 cases (selection in line 4). The lines 2 and 3 are used only for P4 case, line 5 for both cases, while lines 7-9 are used only for scenario case. The last 3 lines are common for both cases. The “postscript base filename” is used to form the names of output files. When this file is generated automatically in classification related programs the option of not printing site symbols is selected and the group number is printed centred at the real coordinates. The option of printing site symbols is useful for delimitation of a contour, but not good for final graphs.

In this script there are used 3 auxiliary programs:

cambiaXY – It changes the format of the file with coordinates (x,y) of the position of sites to the required by GMT program “pstext”

cambiaNaN – It changes the value assigned to the grid points without information from default “NaN” to the one used by SURFER for Windows program: 1.70141+e38

estadigra – It calculates the general statistics for Z component of input files (extremes, average, median, absolute first moment, standard deviation, variance, skewness and kurtosis)

General procedure for scenario studies:

- Run 'mapesce' for for selecting GM data from 'profile.correl.zip' file and to add coordinates of the sites for each profile. It will place the results in 3 directories GMrat, HVrat and 2D in dependence of the kind of data. Process 2D directory only for equivalent earthquakes. Do not join profiles calculated with earthquakes of different magnitudes or at different distances.
- Join the files of the same kind for profiles using the program “joinmap”. It not only join the files, but creates the input data for gridding and plotting all the GM data and a script (joinmap.job) for do this process (type sh joinmap.job)

c) Parametric and non-parametric correlation analysis

In a run of a profile by job2D a parametric and non-parametric correlation analysis is done for all the variables contained in the files with 2D/1D ratio of GM parameters. The result of this analysis is printed in the files “*.cor” included in each “profile.correl.zip” compressed file (see annex 2). The name of the variables in file “ground7.rat.cor” are not printed correctly and has to be corrected by running program “cambiaTAB”.
In each file you will see the values of correlation coefficients with confidence and other related measures, obtained by different methods. You have to use this information in the following way:

1. Determine which variables are statistically independent and drop from analysis all the redundant variables. For example, it is common than Amax, Vmax, Dmax and I_{arias} be strongly correlated. In that case, you should use only one of them; select the one that you consider that will be more useful for the final users of the microzoning. It means that, at the end of this analysis from the maps that you have plotted before it has sense to use only a fraction
2. Perform this analysis for all the components (radial, transverse and vertical) and search for the compatibility of all the results between them. Then add the results of the analysis of the H/V ratios and and you will have the selection for one profile.
3. Repeat the analysis for all the profiles and search for the compatibility of all of them. Then you will have a selection of GM parameters to be used in seismic microzoning.

Although the correlation analysis is done for single GM parameters, their results may be useful also for selecting which functions f(x,f) that has been processed for classification will be useful for seismic microzoning.
There are some programs that may help in the preparation of the data for the analysis:

Program cambiaTAB

It changes the names of the variables in the file ground7.rat.cor that are written incorrectly in the original file. The output is the same “ground7.rat.cor” file. Run always this program once you have unzipped the file “profile.correl.zip”.

Program uneTABcomp joins all the correlation analysis results for a profile in a single file

This program is called by 'uneTABpcomp “name of profile”' in the same directory where you unzipped the file “profile.correl.zip”. It searches for all the “.cor” files and joins them table by table. The result is a file named “profile.rat.cor”. The use of this program allows to make together the items 1 and 2 of the paragraph “Parametric ...”. It is very important to run program cambiaTAB before.

Program uneTABperf joins the correlation analysis results for all the profiles in a single file

In the input file “une.inp” you have to put the list of files “profile.rat.cor” for all the profiles. It joins those files table by table. The result is a file named “totalTAB.cor”. The use of this program allows to make together all the items of the paragraph “Parametric ...”, but the resulting file is very big.

Program borraVAR deletes selected variables from files with correlation results

It deletes variables from files of the format “profile.rat.cor” and “totalTAB.cor”, that are specified in the input “borraVAR.inp:

```
name of input file”
totalTAB.cor
variables to delete:
r.Dmax_ra
t.Dmax_ra
z.Dmax_ra
```

The output is a file of the same format, which name is formed by adding a “.mod” code before the “.cor”, that at the end has a list of deleted variables. This program may help in the analysis of the correlation results.

d) automatic run of the most of the programs

Although you probably don't need to generate all the possibles files and plots that these programsprocesso do, the run of them (with the exception of P4) takes a little time, and its is convenient to generate the maximum possible in an automatic way. Then you can select the ones that you will use for microzoning.

“microMapP4” is a program that generates scripts for automatic run of classification, 2D plotting proceses and correlation analysis

The process of scenario plotting requires little intervention. You have only to select the proper values for the parameters of the “fdtriangul.par” file. Then, they are passed by input file joinmap.par and program joinmap creates the particular input file for each GM variable. This is not the case of the classification, when you have to interact with program P4, and the decisions should be different from one GM function to other. Then the automation of the first case can be done completely, while for the second case only partially.
This is done by program microMapP4.For using this program you have to:

- place in a directory the files profile.correl.zip and profile.dat_P4.zip for all the profiles
- create the files joinmap.par (program joinmap) and join_ma.inp (program join_ma) that will control the process and put them in the same directory
- create the input files mapesce.par.profile (program mapesce) and cu2ma.par.profile (program cu2ma) for all the profiles given in joinmap.par and join_ma.par and put them in the same directory
- place in this directory also the file with contour to plotted inside the

working area

The program reads the files joinmap.par and join_ma.par (may not be equal, you can do the selection of profiles that you want), then it checks for the presence of files mapesce.par.profile and cu2ma.par.profile. If they are present, program creates 3 scripts for processing the data:

```
cre_mapsES – for 2D scenario
cre_mapsP4 – for classification with P4
creaMaps   – both
```

By running cre_mapsP4 you will obtain the input matrices for classification with P4. In the directory P4/ma it places the single matrices by profile while in the directory P4/matP4 it places the joined ones. In the directories P4/profile/dat_p4/ps it places the plotting of input not sampled curves with the sampled ones.

By running cre_mapsES you will obtain the input files for GMT plotting of a 2D scenario and the corresponding plots and data files for each joined GM data file in directories correl/esc2D, correl/escGM and correl/escHV. Take into account that all the files “code.XY.ps” are equal, because they correspond to site distribution, that is the same for all parameters.

All these scripts check before for the existence of the directories "P4" and correl". If they exist are deleted. ***Be careful if you want to do several runs changing parameters.*** In that case you need to change the name of the directories before each new run.

WARNING: *Take into account that the plots placed in directory correl/esc2D correspond to GM data extracted from 2D seismograms only. They have sense only in the case that all profiles correspond to the same scenario, i.e., the same earthquake at the same distance.*

Additionally the scripts run in each profile the programs “cambiaTAB” and “uneTABcomp”, place the results for all profiles in the directory correl/cor and then run the program “uneTABperf”. Below is the list of a directory contain for a complete run of microMapP4 with 3 profiles (M11,M12,M13):

a) Initial data for synthetic seismograms' calculation: GM parameters (“correl”) and functions (“dat_P4”)

```
M11.correl.zip
M11.dat_P4.zip
M12.correl.zip
M12.dat_P4.zip
M13.correl.zip
M13.dat_P4.zip
```

b) Data for program “cu2ma”

```
cu2ma.inp.M11
cu2ma.inp.M12
cu2ma.inp.M13
```

c) Data for program “mapesce”

```
mapesce.par.M11
mapesce.par.M12
mapesce.par.M13
```

d) Parameter files for programs “join_ma” and “joinmap”

```
join_ma.inp
joinmap.par
```

e) Base contour to be included in maps

```
manzanillo.xy
```

e) Microzoning

Seismic microzoning is a very complex task. Bellow will be described some procedures that were tested during SMLAC project.

e.1) Initial considerations

At this stage you can have, for each profile, a maximum of:

- maps for 24 2D GM scenario (with corresponding files for gridding with 3 different methods)
- maps for 56 GM 2D/1D ratios (with corresponding files for gridding with 3 different methods)
- maps for 24 H/V relations 2D/1D ratios (with corresponding files for gridding with 3 different methods)
- matrices and result of classification for 24 functions f(x,f)
- a parametric and non parametric correlation analysis of all combinations of GM ratios

The microzoning should be done with this information, and it is important to select which of them are really significant. There are different procedures that can give good results wit different datasets.

- The use of the results of classification of GM function, represented on maps, and microzones drawn by eye fitting with the help of geological-geophysical information.
- The use of maps prepared with ground motion parameters, and microzoning drown with the help of a GIS.
- The use of maps prepared with ground motion parameters, and microzoning made by processing them with P4 program.

In all cases the procedure of delineating zones is user dependent. As a result you should have a map with microzones. In particular it will be explained the third option.

e.2) Microzoning with the help of P4 program

The maps prepared with GM parameters, as their are the result of a interpolation process, cover all the study area in a regular grid; then, for each point in the grid can be defined a “vector”, which components correspond to the values of the GM parameters in this point. Normally, all the GM parameters will not have similar behaviour, then you have to select some of them that presents in the maps somewhat similar characteristics. Once selected, pick up the corresponding xyz files (if you ran “crea_mapsES” or “creaMaps” they will be placed in correl/GMrat and correl/HVrat). You will need also a base contour file (for plotting maps) and over it you have to draw and digitize a blanking contour for cleaning your input xyz files from not desired or less reliable zones.
The processing of data is as follows:

1. Select the xyx files to use.
2. Use program “xyz2ma” to create input matrices (one with real data and other with integer data – order number of interval in z.
3. Select one of these matrices and use program “P4” for non supervised classification of input data. Save the results in the .htx file.
4. Use plotclass1 for making the graphs and map (name_of_matrix.P4.ps) corresponding to this solution. The results will appear in a directory called after the name of input matrix.

5. Move to this directory and check results. If you don't want additional regrouping use program htx_contour (see below) for plotting a “step” kind of zoning map (name_of_matrix.P4.ps) .
6. If you want to regroup your classification analyse the dendrogram of “P4” together with “.htx” for taking the decision about groups to join. Then run “joingroup1”. The results (including a new “step” kind of zoning map) will be placed in a directory named “join”
7. Analyse these results. If you are satisfied you have to smooth the “step” map using any drawing program. It can be very helping to this process if you use as base map in “plotclas1.inp” a one that contain not only cartographic details, but geological and geophysical information.

These is a “try and error” procedure, and probably you will need to repeat some parts or the whole process several times until you will be entirely satisfied.

program “xyz2ma” process the '.xyz' files and creates input matrices for P4

This program reads the input files and check for compatibility of grid points between them, blanks the input data and divides in a user given number of intervals the z range. It has as result two matrices for P4 input, one with real z values and a similarity criterion
"<given_value=(zmax-zmin)/number_of_intervals",
and other with integer values (order number of intervals in z) and similarity criterion "equality".
It uses the input file xyz2ma.inp. Due to limitations in the length of variable names in P4, the names of files put as variable names are cut (neglected characters until first '.', and last 7). It uses the input file xyz2ma.inp:

```
generic name for output file (below):
prueba
7   number of intervals to discretize input data
4   number of files with xyz data
1   blank outside a contour (1 - yes, 0 - no)
name of contour file below (single closed, first line not read):
blanqueo.bln
list files with xyz data below:
GMrat.r_AEI_ra.MC.xyz
GMrat.r_EPA_ra.MC.xyz
GMrat.z_EPA_ra.MC.xyz
GMrat.z_RSRmax_A.MC.xyz
```

This program can be used in two different ways. The first is for additional processing the results of the application of different algorithms for gridding and contouring the point data obtained in GM processing, as represented in the input file above presented. As can be seen, these files are the result of the gridding process for one particular method (minimum curvature). The use of a blanking contour should be decided taking into account the reliability of results of the selected gridding process.
The other use is for classifying with P4, not the data of functions of frequency as already seen, but the GM data before doing the process of gridding and contouring. In this case we bring the data for the GM parameters in the original form. In this case it is not necessary to use a blanking contour because all data are equally reliable.

```
generic name for output file (below):
prueba
4   number of intervals to discretize input data
4   number of files with xyz data
0   blank outside a contour (1 - yes, 0 - no)
name of contour file below (single closed, first line not read):
blanqueo.bln
list files with xyz data below:
allGMrat.r_RSRmax_V.dat
allGMrat.r_T_EIRmax.dat
allHVrat.f_RsVmaxSA.dat
allHVrat.f_TsVmaPV.dat
```

With difference with the case of P4 matrices for functions of frequencies, when curves to be classified has a real physical meaning, in the case presented here with are not treating with functions, but with "functionals", then we have defined a measure that allows us to process them with P4. Only the matrix for intervals has the possibility of been plotted. Program plots this matrix (throgh gnuplot) with a name formed by the generic one given in input plus information about number of intervals used in z range.

“htx_contour” is a program for contouring post data in a plot

This program takes the information created in a post file by programs “plotclas1” or “joingroup1” (a “xyz” file with z – integer) and draws a contour between grids with different z-values. It assumes that the grid is complete. If you use the program with grids not satisfying this condition the results will be wrong. If you want to process not equally-spaced data or incomplete grids, you need to generate a complete grid with any interpolating program. The data that we propose to use in this methodology are the GM sets interpolated by different programs of GMT. The program draws online a map with a name formed by “name_of_matrix.ps” with perl script “fdpostmap.pl” that makes use of GMT programs.

It reads the input file “htxco.inp”:

```
Output bln file:
prueba_D.xy
1                  output bln format (0 -SURFER; 1 -GMT)
Input files (post and contour to add)("none" for no adding contour):
prueba_DMpost.dat
bayamo.xy
1                  input bln format (0 -SURFER; 1 -GMT)
Plotting information:
 516777.80    minimum longitude (x)
 525538.20    maximum longitude
 187669.45    minimum latitude   (y)
 194743.55    maximum latitude
 5    dimension of posts (always integer)
```

Program “plotclas1” creates a typical one from its input data. In case of problems in obtaining the final map, you have to check it (take special care about the format of contour file).

Warning: Program “plotclas1” creates a input file “htxco.inp” and technically you can run the program. In the same way, program “joinmap1” runs “htx_contour” if exist “htxco.inp”. But this program is only FOR REGULAR GRIDDED DATA; you can not use its results in other cases.

e.3) General recommendations for microzoning:

- Use the results of correlation analysis for discriminating among the variables to be selected
- Extend these results to the selection of functions f(x,f) to be classified with P4
- Plot the results of classification of functions
- Combine the selected plots of variables and results of classification of functions with geological-tectonical-structural information of the study area for delineating zones with quasi homogeneous behaviour of ground motion
- Document each created zone with the must of average information about ground motion parameters and functions from the sites located inside it using program par_micro

“par-micro” prepares the "documentation" for a seismic microzoning

For documentation are understood the tables with average values of selected GM parameters and graphics with average of selected GM functions for all the sites placed inside each microzone. Previous to its use you have to:

- Prepare a map of delimited microzones in SURFER's "bln" or GMT's xy format. Each zone may be separated in different closed contours. In cases of zones included in others, the external ones should be subdivided in at least two closed contours in order that algorithm of the program works properly
- Select the parameters of GM ratios that should be averaged in each microzone
- Select the functions of GM ratios that should be averaged in each microzone

Input files: main "micro.inp" with: names of "bln" file with contours, GM ratios input files, GM ratios functions input files and indication about actual number of microzones corresponding to each contour in "bln" file

```
fichero de contornos:
mapa.bln
0        (0 - SURFER bln, 1 - GMT xy)
ficheros de GM:
3
allGMrat.r_AEI_ra.dat
allGMrat.r_Amax_ra.dat
allGMrat.r_EPA_ra.dat
5
JOIN_hvd2d1d.ma
JOIN_hvpv2d1d.ma
JOIN_hvv2d1d.ma
JOIN_r2d_rsra.ma
JOIN_r2d_rsrd.ma
renumbering of zones in contour file (joining or simple reordering)
1              (0 - no, 1 - yes)
if "yes" put one line for each contour of .bln with new number
1
2
2
2
3
```

There is a point that requires additional explanation. These is the part of the input file after “renumbering of ...”. Sometimes you can not describe a zone by a single contour; for example, two isolated zones of the same category. But there is a limitation due to the subroutines that determines if a point is inside a zone. You can not give to program a zone that is inside another one. When this is the case, the embeded zone have to be subdivided in several parts so the embedded zone be always outside of the new ones. For example, in the case of two concentric circles, the inner one remains as a zone, but the space between the two circles should be subdivided in at least two sectors. Then, you have to inform to the program to which zone in the map corresponds each one of the contours in the input contour file.

Output files:

- micro.res -- a wide report of process plus results
- micro.tab -- a table with GM ratios averaged for each microzone
- micro*.xy -- files that contain average curves of GM ratios for one parameter and one microzone in 3 columns: (frequency, period, GMratios). These 2D/1D GM ratios are considered as transfer functions rock-soil
- micro*.idl -- files with data to be plotted with IDL program "pclamu" with average curves of GM ratios for one parameter and all microzones
- pcmu.inp -- input parameter file for IDL program "pclamu" that uses the files micro*.idl

To complete the process you need to run "pclamu" (idl-->pclamu-->exit) after this program, and you will obtain graphics of the averaged functions for all microzones.

f) Preparing data for seismic codes

f.1) Using transfer functions (GM ratios functions) for scaling probabilistic functions in rock to obtain soil ones.

It is very extended the practice of calculating response spectra for rock in a PSHA. There are reliable attenuation relations that can be used in this case. Nevertheless, it is not the situation for soil. The attenuation relations are obtained in that case for an average soil and in general seem to be underestimated with respect to which it is expected. The use of an estimation of a PSHA response spectrum for rock and a transfer function rock-soil calculated for a real soil in a microzoning work is a more reliable way of obtaining an estimation of response spectrum for soil. It is valid for any other function of frequency that can be obtained for rock and the corresponding transfer function rock-soil.

"compespec" calculates the response spectra in soil from response spectra in rock and transfer function

This program reads the response spectrum in rock and transfer function and another spectrum for comparison, adjust sampling of both functions, calculates soil spectra as rock*transfer_function and plots rock response spectrum, soil response spectrum, transfer function and response spectrum for comparison. It allows to process together several combination of data. It uses a control file

named "compspec.inp" and the corresponding spectra and transfer functions files:

```
Ficheros de función de transferencia
3     numero de funciones de transferencia
2  1  columna donde estan el periodo o frecuencia y orden (0 - normal, 1- inverso)
3     columna donde esta la función de transferencia
3     numero total de columnas
1     tipo de espaciamiento de los puntos (numero de orden)
micro_JOIN_r2d_rsra.ma_zona_02.xy
micro_JOIN_t2d_rsra.ma_zona_02.xy
micro_JOIN_z2d_rsra.ma_zona_02.xy
Ficheros base
2      numero de espectros base a multiplicar por función de transferencia
1  0   columna donde esta el periodo o frecuencia y orden (0 -normal, 1 -inverso)
4      columna donde está la función base
6      numero total de columnas
2      tipo de espaciamiento de los puntos (numero de orden)
trenRigido.txt
trenRoca.txt
Ficheros de espectros a comparar
1      numero de espectros en suelo a comparar con base*Ftrans
1  0   columna donde esta el periodo o frecuencia y orden (0 -normal, 1 -inverso)
4      columna donde está la función base
6      numero total de columnas
2      tipo de espaciamiento de los puntos (numero de orden)
trenBlando.txt
Decisiones tomadas
1     tipo de espaciamiento de los puntos preferido
1.0   valor minimo de x (poner numero grande negativo para usar el de los datos)
1000  valor maximo de x (poner numero grande positivo para usar el de los datos)
1     tipo de grafico (0-bilinear, 1-bilog)
```

The sampling adjustment is made in the following way. For each kind of input file (3 possibles) you assign an order number for its sampling. If it is the same as a previous one the number is repeated. You have to select the kind of sampling for your results.

The output is formed by three files for each combination rock_spectrum-transfer_function-comparison_spectrum: one ".txt" with a table of the values to be plotted (period, rock spectrum, transfer function, comparison spectrum, calculated soil spectrum) that is used for "gnuplot" as data for plotting (a ".ps" graph) and a ".pdf" version of this graph, created by "ps2pdf" after running of "compespec.job". The plotting and conversion by other programs is done during the program run.

Program par-micro prepare the input file of transfer functions. The "external" files with rock and comparison functions should be normal ASCII tables with one column of period or frequency and other with values of spectrum. This file can have several columns (maximum 10) and are selected the necessary.

Although the explanation talks about response spectrum, this process can be applied to any kind of GM function of frequency (or period). The program "compespec" can be easily modified as new needs appear.

Annex 1. Files with information about frequency dependent ground motion parameters [functions (f,x)]

profile.hva2d1d – acceleration H/V ratio 2D/1D
profile.hvd2d1d – displacement H/V ratio 2D/1D
profile.hvpv2d1d – pseudo velocity H/V ratio 2D/1D
profile.hvv2d1d – velocity H/V ratio 2D/1D
profile.r2d.rsra – radial response spectral ratio 2D/1D for acceleration
profile.r2d.rsrd – radial response spectral ratio 2D/1D for displacement
profile.r2d.rsrpv – radial response spectral ratio 2D/1D for pseudo velocity
profile.r2d.rsrv – radial response spectral ratio 2D/1D for velocity
profile.t2d.rsra – transverse response spectral ratio 2D/1D for acceleration
profile.t2d.rsrd – transverse response spectral ratio 2D/1D for displacement
profile.t2d.rsrpv – transverse response spectral ratio 2D/1D for pseudo velocity
profile.t2d.rsrv – transverse response spectral ratio 2D/1D for velocity
profile.z2d.rsra – vertical response spectral ratio 2D/1D for acceleration
profile.z2d.rsrd – vertical response spectral ratio 2D/1D for displacement
profile.z2d.rsrpv – vertical response spectral ratio 2D/1D for pseudo velocity
profile.z2d.rsrv – vertical response spectral ratio 2D/1D for velocity
profile.rad2d.rsp – 2D radial response spectra
profile.tra2d.rsp – 2D transverse response spectra
profile.ver2d.rsp – 2D vertical response spectra
profilef0.rac.eir – radial energy input ratio 2D/1D scaled
profilef0.rac.eir_lg – logarithm of radial energy input ratio 2D/1D scaled
profilef0.tac.eir – transverse energy input ratio 2D/1D scaled
profilef0.tac.eir_lg – logarithm of transverse energy input ratio 2D/1D scaled
profilef0.zac.eir – vertical energy input ratio 2D/1D scaled
profilef0.zac.eir_lg – logarithm of vertical energy input ratio 2D/1D scaled

'***profile***' means the name of the problem that you give in the line 24 of pfdg9.par and '***f0***' means that the signals used in calculations of the parameters are scaled to the magnitude given in line 20 of pfdg9.par

Annex 2. Files with information about ground motion parameters [functions (x)]

profile.esce – GM parameters of the 2D synthetic seismograms (see annex 4)
profile.r.rat – ratios 2D/1D for GM data and response spectra for radial component (see annex 5)
profile.t.rat – GM ratios 2D/1D for GM data and response spectra for transverse component
profile.z.rat – GM ratios 2D/1D for GM data and response spectra for vertical component
ground7.rat – ratios 2D/1D for H/V parameters for all components (see annex 6)
profile.r.rat.cor – correlation and non parametric correlation analysis for the variables included in file “profile.r.rat”
profile.t.rat.cor – correlation and non parametric correlation analysis for the variables included in file “profile.t.rat”
profile.z.rat.cor – correlation and non parametric correlation analysis for the variables included in file “profile.z.rat”
ground7.rat.cor – correlation and non parametric correlation analysis for the variables included in file “ground7.rat”

'***profile***' means the name of the problem that you give in the line 24 of pfdg9.par

Annex 3. Short description of program “P4”

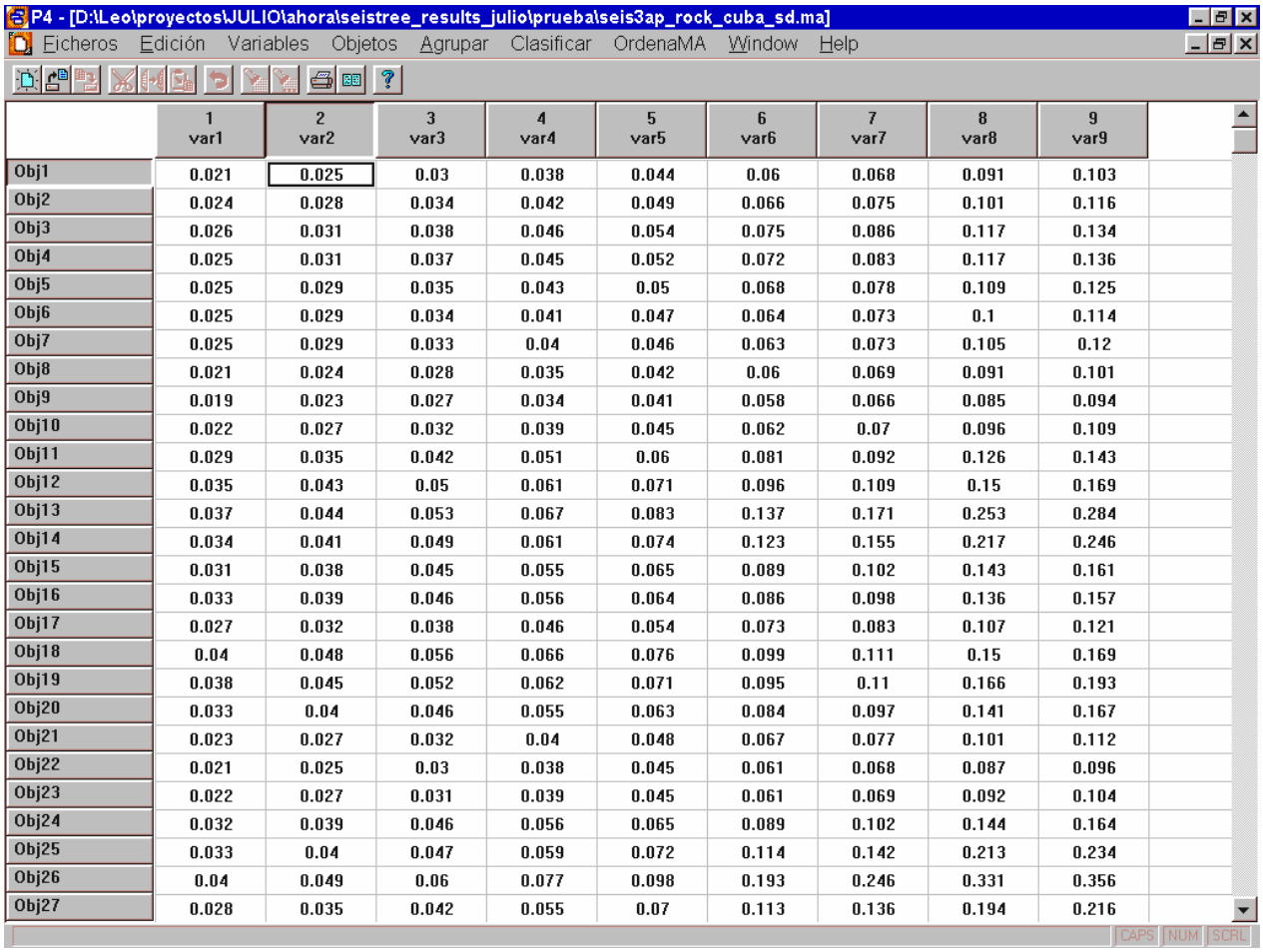
a) General considerations

First of all you have to open (option “Ficheros”) an initial matrix that can be a learning matrix for supervised classification or a data matrix for non supervised classification. Once a matrix is opened (Fig. 1) all the windows in the main menu are activated, and between them the following:

- **Agrupar**: non supervised classification (the one to be used in this process)
- **Clasificar**: supervised classification
- **OrdenarMA**: ordering of the objects in the matrix according to the values of the variables with more dispersion

By clicking two times in the buttons with “Objn”, with n = number of the object and “varn”, with n = number of the variable it is opened a window that allows to edit the corresponding object or variable. In our case “object” is a site and “variable” is the frequency.

WARNING: The option of editing do not work well when P4 is ran in Linux with “wine”. It writes incorrect results.



	1 var1	2 var2	3 var3	4 var4	5 var5	6 var6	7 var7	8 var8	9 var9
Obj1	0.021	0.025	0.03	0.038	0.044	0.06	0.068	0.091	0.103
Obj2	0.024	0.028	0.034	0.042	0.049	0.066	0.075	0.101	0.116
Obj3	0.026	0.031	0.038	0.046	0.054	0.075	0.086	0.117	0.134
Obj4	0.025	0.031	0.037	0.045	0.052	0.072	0.083	0.117	0.136
Obj5	0.025	0.029	0.035	0.043	0.05	0.068	0.078	0.109	0.125
Obj6	0.025	0.029	0.034	0.041	0.047	0.064	0.073	0.1	0.114
Obj7	0.025	0.029	0.033	0.04	0.046	0.063	0.073	0.105	0.12
Obj8	0.021	0.024	0.028	0.035	0.042	0.06	0.069	0.091	0.101
Obj9	0.019	0.023	0.027	0.034	0.041	0.058	0.066	0.085	0.094
Obj10	0.022	0.027	0.032	0.039	0.045	0.062	0.07	0.096	0.109
Obj11	0.029	0.035	0.042	0.051	0.06	0.081	0.092	0.126	0.143
Obj12	0.035	0.043	0.05	0.061	0.071	0.096	0.109	0.15	0.169
Obj13	0.037	0.044	0.053	0.067	0.083	0.137	0.171	0.253	0.284
Obj14	0.034	0.041	0.049	0.061	0.074	0.123	0.155	0.217	0.246
Obj15	0.031	0.038	0.045	0.055	0.065	0.089	0.102	0.143	0.161
Obj16	0.033	0.039	0.046	0.056	0.064	0.086	0.098	0.136	0.157
Obj17	0.027	0.032	0.038	0.046	0.054	0.073	0.083	0.107	0.121
Obj18	0.04	0.048	0.056	0.066	0.076	0.099	0.111	0.15	0.169
Obj19	0.038	0.045	0.052	0.062	0.071	0.095	0.11	0.166	0.193
Obj20	0.033	0.04	0.046	0.055	0.063	0.084	0.097	0.141	0.167
Obj21	0.023	0.027	0.032	0.04	0.048	0.067	0.077	0.101	0.112
Obj22	0.021	0.025	0.03	0.038	0.045	0.061	0.068	0.087	0.096
Obj23	0.022	0.027	0.031	0.039	0.045	0.061	0.069	0.092	0.104
Obj24	0.032	0.039	0.046	0.056	0.065	0.089	0.102	0.144	0.164
Obj25	0.033	0.04	0.047	0.059	0.072	0.114	0.142	0.213	0.234
Obj26	0.04	0.049	0.06	0.077	0.098	0.193	0.246	0.331	0.356
Obj27	0.028	0.035	0.042	0.055	0.07	0.113	0.136	0.194	0.216

Fig. 1. Main window

When it is used the option “Clasificar → Dendrograma → Todos”, it appears a window with the dendrogram (it may require some time in dependence of the number of objects-sites to classify), with a lower button that allows to select

the desired option (“componentes conexas” - connected components, “conjuntos compactos” - compact sets and “conjuntos compactos maximales” - maximal compact sets) (Fig. 2). Over this dendrogram is selected the level β_o, that corresponds to a horizontal line. The dendrogramas corresponding to compact sets (Fig. 2) and connected components (Fig. 3) are different. For the case of maximal compact sets, it will appear the dendrogram of compact sets, which is used to select the β_o, because this kind of grouping do not allow to draw a specific dendrogram.
In these windows it aperas a button named “Particion”, which allows to execute another actions as to check grouping results, save them, and in the particular case of maximal compact sets, to create a learning matrix by using the average curve of each one of the formed groups (Fig. 4).
By pressing the button “Guardar” a common saving window is opened. By pressing button “Salir” you return to the window behing and repeat classification process with other decisions.

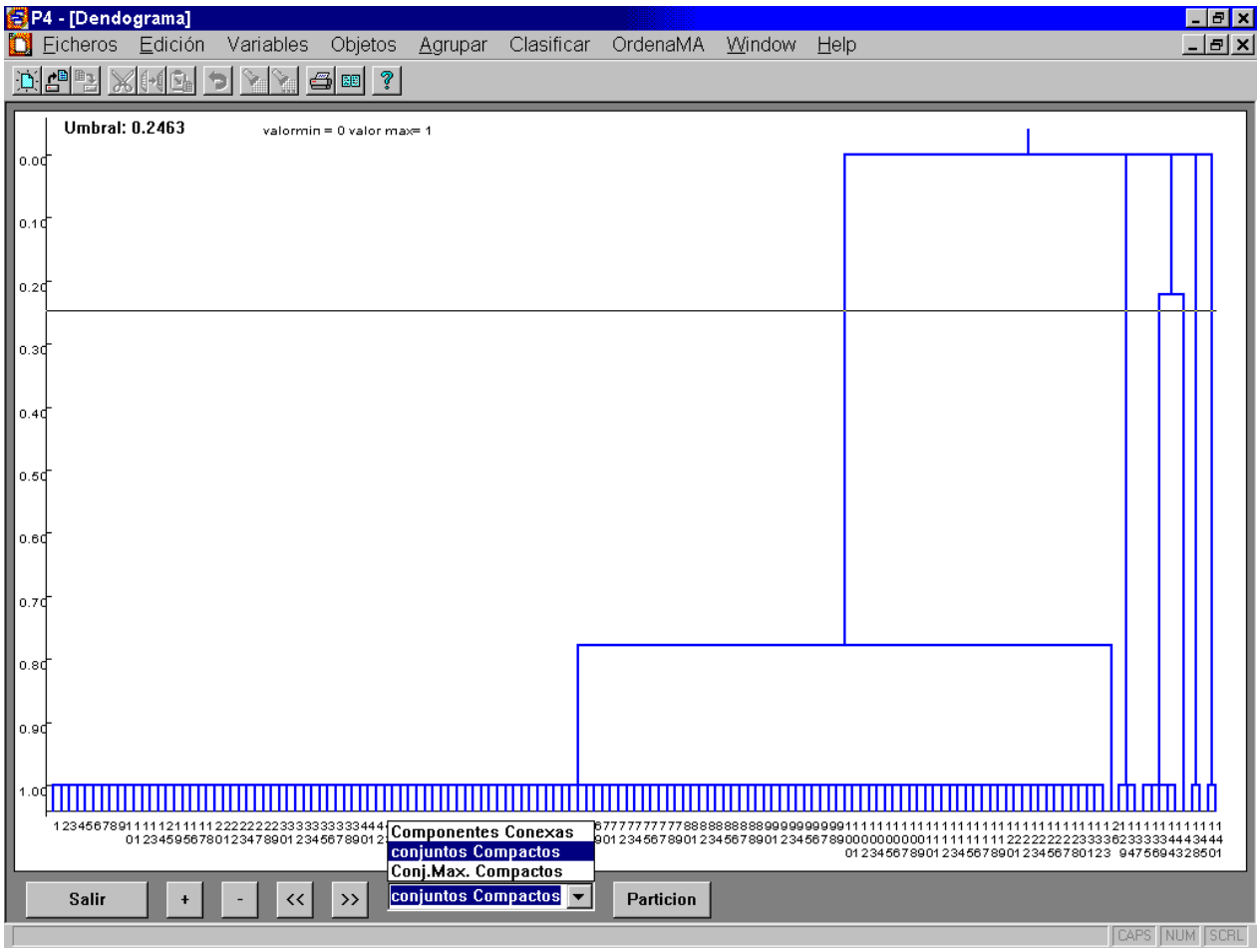


Fig. 2. Grouping by compact sets. The options for grouping are shown. The horizontal line in the plot corresponds to the level β_o. In the top-left corner of the plot it its shown this value.

Note: When you use program “cu2ma” for creating the initial matrices, it is selected the comparison criteria A-2 (see below) with $\varepsilon_t = \{max(x_i) - min(x_i)\}/10$.

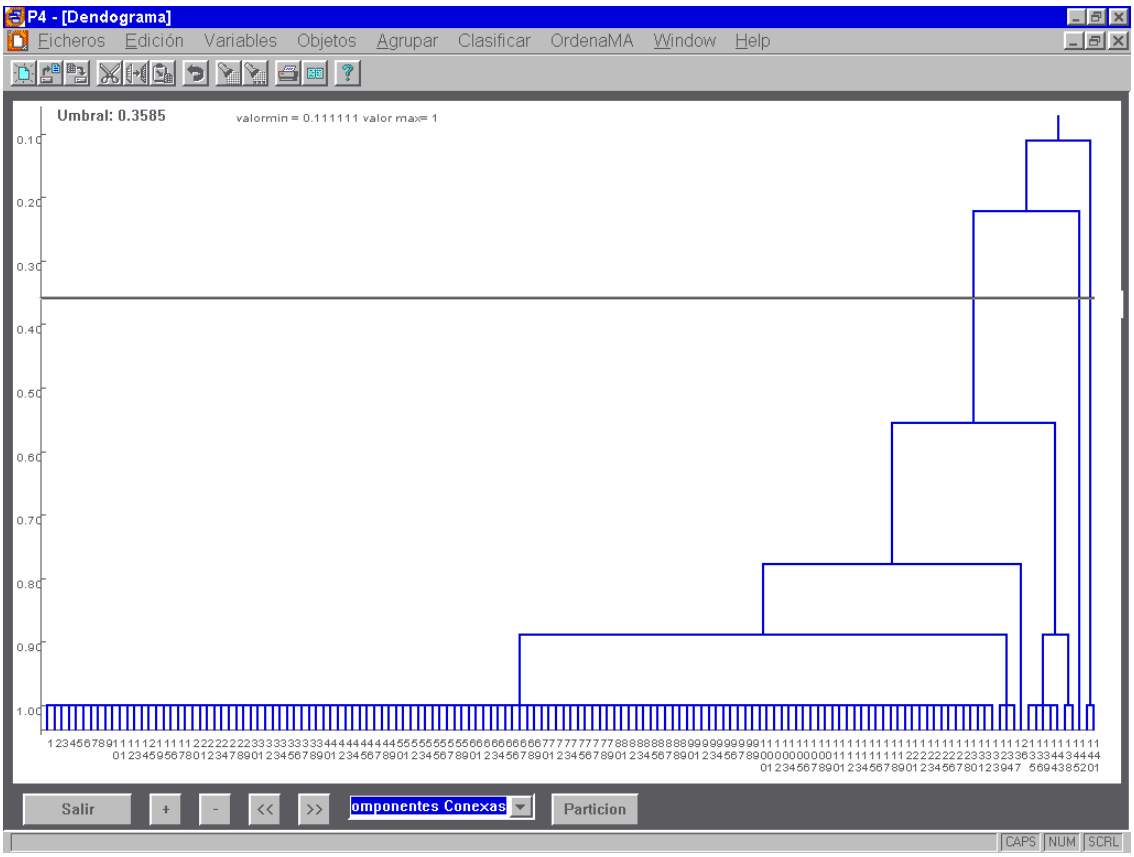


Fig. 3. Grouping by connected components

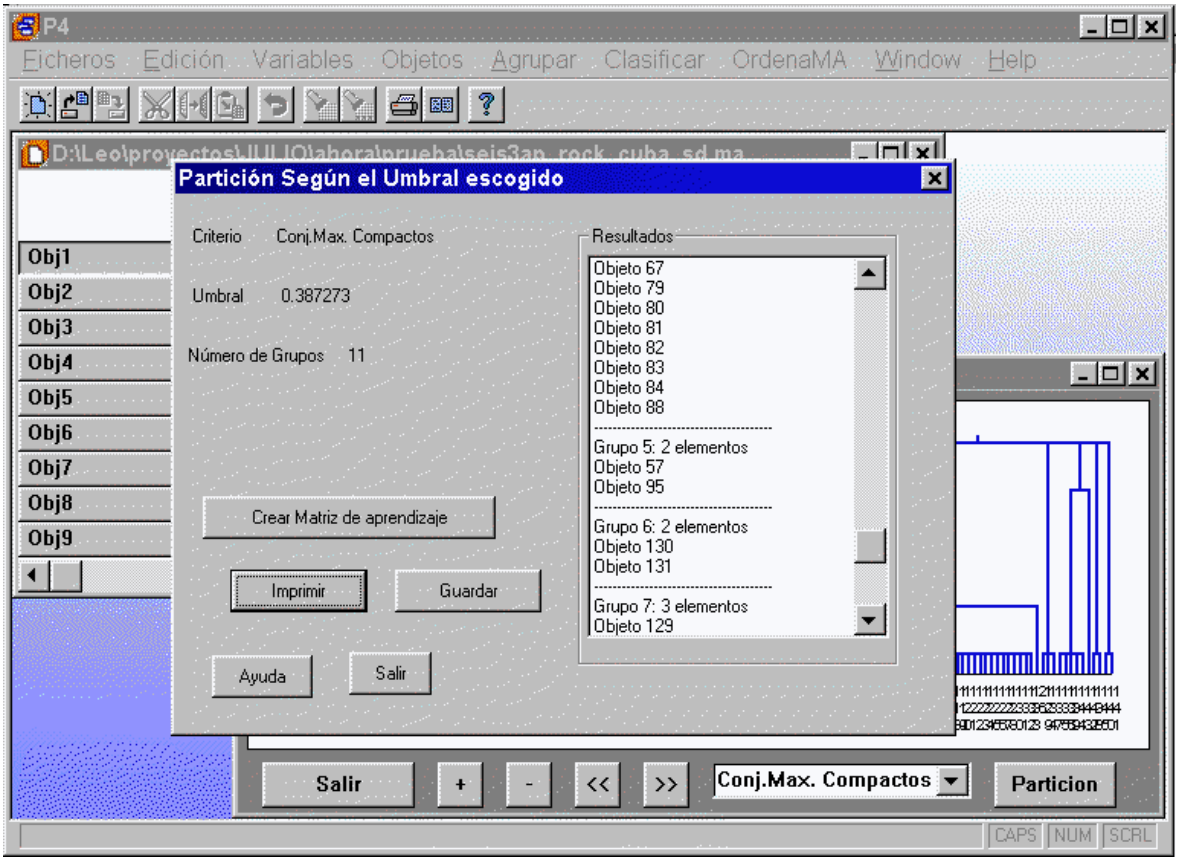


Fig. 4. Results of grouping by the maximal compact sets algorithm. They are shown after clicking on the button “Particion” (right-bottom of the window behind). For each group are shown the group number, the number of elements inside it, and the elements numbers (Objeto ##).

b) Classification by logical-combinatorial algorithms

The main theoretical principles of the PROGNOSIS system (Ruiz et al.,) is :

Let $\Omega_A = \{\Omega_1, \ldots, \Omega_k\}$. A (total) similarity function is an operator:

. $$\Gamma : \prod_{S=1}^{k} Q_S \to L\ ,\ L \subseteq [0,1] \qquad \text{(A-1)}$$

PROGNOSIS system has as default the following similarity function:

$$\Gamma(0_i, 0_j) = \frac{\gamma(0_{ij})}{K} \sum_{S=1}^{k} \rho(\Omega_S)\, \varpi_S\, \beta_{ij} \qquad \text{(A-2)}$$

with is a weighting parameter associated to the comparison between objects Oi and Oj; in the same way is a weighting parameter associated to the support set and could be a function of the informational weight of features that appear in such support set, for example

$$\rho(\Omega_S) = \frac{\sum_{x \in \Omega_S} \rho(x)}{|\Omega_S|} \qquad \text{(A-3)}$$

where ρ(x) is the informational weight of feature x.

Program P4 is a particular realization of this system. The selected algorithms for present application start from the general conditions:

- Let the curve corresponding to the point "*j*" be the object "O_j", and the value of the ordinate at frequency number "*i*" be the variable "x_i". Then, the value of ordinate at the frequency number "*i*" in the curve number "*j*" will be "$x_i(O_j)$". In the case of the table of ground motion parameters or the microzoning data the object O_j is the same, and the variables contain the values of the ground motion in each column of the table or the coding of the results of a previous classification.
- Let $S(O_i,O_j)$ be the similarity function between objects O_i and O_j (to be defined below). Two objects O_i and O_j are β_o-*similar*, if and only if $S(O_i,O_j) \geq \beta_0$ where β_0, the level of the classification, is between 0 and 1.
- An object belongs to a connected component (Voronin et al., 1968) if all its β_0 - *similar* objects belong to this set.
- An object belongs to a compact set (Sirotinskaya. 1986) if the most β_0 - *similar* to it is into this set too, or if it is the most similar to other object belonging to the set.
- An object belongs to a maximal set if it is β_0 - *similar* to all the objects belonging to the set.
- An object belongs to a maximal compact set if the most β_0 - *similar* to it is into this set too, or if it is the most similar to other object belonging to the set and at the same time all the elements of the set are mutually β_0 - *similar*.
- The compact sets or the connected components are graphically represented in a dendrogram, where the different levels represent the β_0 - similarity, which can be used to build groups. Selecting interactively, over this scheme, a level β_0, a particular partition in β_0 - compact sets or in β_0 - connected components can be determined

(Pico, 1999).

- A unique tree forms the dendrogram for connected components, i.e., there is β_0 level such that all the objects are connected. Besides, the dendrogram for compact sets is, in general, formed by several isolated trees; i.e., while reducing the level β_0, the number of independent groups cannot be reduced under certain limit. This behaviour is due to the fact that in the case of compact sets, the condition of connectivity between objects is more restricted than in the case of connected components.
- The maximal compact set is obtained by finding maximal sets from a partition of compact sets.
- An object cans belong simultaneously to several maximal compact sets. This characteristic impedes their representation in a dendrogram.
- In order to avoid the classification of an object in several different groups, it is necessary to introduce additional criteria in the algorithm. One of the simplest decisions is that once an object is assigned to a group, it will not be considered any more. Under this condition can be obtained several different classifications, depending on the order in which the objects are compared. To get some uniformity in this process, the objects are ordered in descending order taking into account the variance of their component variables. The algorithm with these additional criteria will be called "maximal compact sets modified".
- The procedure starts by determining the main β_0 - compact sets or β_0 - connected components. For the maximal compact sets, the β_0 is fixed from the compact sets dendrogram and a maximal compact sets partition is calculated. Finally, an average curve for each final set is calculated.
- The comparison criterion of a variable x_t is a function C_t that can be defined in different ways, according to the kind of variable, domain of definition, etc. Let us see three possible variants:
  1. Let $max(x_i)$ and $min(x_i)$ be the extremes of the variable x_i over all the objects. Then:

$$C_t\left[x_t(O_i), x_t(O_j)\right] = 1 - \frac{|x_t(O_i) - x_t(O_j)|}{\max(x_t) - \min(x_t)} \qquad \text{(A-4)}$$

  2. Let ε_t be a threshold related to the characteristics of the domain of definition of variable x_t :

$$C_t\left[x_t(O_i), x_t(O_j)\right] = \begin{cases} 1 & if \quad |x_t(O_i) - x_t(O_j)| \le \varepsilon_t \\ 0 & \textbf{in other case} \end{cases} \qquad \text{(A-5)}$$

  3. Let variable x_t be formed only by integer numbers (k-valued variable), the comparison is made by simple equality:

$$C_t\left[x_t(O_i), x_t(O_j)\right] = \begin{cases} 1 & if \quad x_t(O_i) = x_t(O_j) \\ 0 & \textbf{in other case} \end{cases} \qquad \text{(A-6)}$$

- The similarity between objects in all the cases is calculated by the

formula:

$$S(O_l, O_k) = \frac{1}{n} \sum_{p=1}^{n} C_t\left[x_t(O_l), x_t(O_k)\right] \qquad \text{(A-7)}$$

where n is the number of considered variables and C_t the criterion above defined.

The initial settings for the classification process depend on the sample characteristics.

In the case of considering the informational weight of the variables, the last formula is transformed in:

$$S(O_i, O_j) = \frac{1}{\sum_{t=1}^{n} \rho_t} \sum_{t=1}^{n} \rho_t C_t\left(x_t(\cdot_i), x_t(0_j)\right) \qquad \text{(A-8)}$$

where ρ_P is the informational weight of the variable P. If the informational weight of the objects is considered also, the formula takes the form (A-2):

$$S(O_i, O_j) = \frac{\gamma_{ij}}{K \sum_{t=1}^{n} \rho_t} \sum_{t=1}^{n} \rho_t C_t\left(x_t(\cdot_i), x_t(\cdot_j)\right) \qquad \text{(A-9)}$$

where γ_{ij} / K should be some kind of normalization for considering the informational weight of the objects.

References for P4 program

PICO, R. (1999): Determinación del umbral de semejanza β_o para los algoritmos de agrupamiento lógico-combinatorios, mediante el dendrograma de un algoritmo jerárquico, SIARP'99, IV Simposio Iberoamericano de Reconocimiento de Patrones. Memorias, pp. 259-265.

SIROTINSKAIA, S.V. (1986): Logical methods for geological information analysis (in Russian), Nedra.

VORONIN, YU. A.; KARATAEVA, G. N.; EPSHTEIN, E. N. (1968): Programs HOLOTYPE for pattern recognition problems solution (in Russian), Alma Atá, 1968.

Annex 4. *GM parameters of the 2D synthetic seismograms*

Amax2_com – maximum acceleration for component com=(tra,rad,ver)
Vmax2_com – maximum velocity for component com=(tra,rad,ver)
Dmax2_com – maximum displacement for component com=(tra,rad,ver)
Aria2_com – Arias intensity for component com=(tra,rad,ver)
AEI_2_com – area under curve of energy integral for component com=(tra,rad,ver)
EIma2_com – maximum of energy integral for component com=(tra,rad,ver)
T_EI2_com – period of this maximum value
EPA_2_com – effective peak acceleration for component com=(tra,rad,ver)
c_SA5%max – maximum response spectra for acceleration at 5% damping for component c=(t,r,v)
c_T_SAmax – period of this maximum value
c_SV5%max – maximum response spectra for velocity at 5% damping for component c=(t,r,v)
c_T_SVmax – period of this maximum value
c_SD5%max – maximum response spectra for displacement at 5% damping for component c=(t,r,v)
c_T_SDmax – period of this maximum value
c_PSSV5%ma – maximum response spectra for pseudo velocity at 5% damping for component c=(t,r,v)
c_T_PSSVma – period of this maximum value

In all this case the codes ***"com"*** and ***"c"*** means that there is a different file for each component.

Annex 5. Ratios 2D/1D for GM data and response spectra

c_Amax_ra – maximum acceleration ratio 2D/1D for component c=(t,r,v)
c_Vmax_ra – maximum velocity ratio 2D/1D for component c=(t,r,v)
c_Dmax_ra – maximum displacement ratio 2D/1D for component c=(t,r,v)
c_I_Ariasra – Arias intensity ratio 2D/1D for component c=(t,r,v)
c_AEI_ra – area under energy integral curve ratio 2D/1D for component c=(t,r,v)
c_EPA_ra – effective peak acceleration ratio 2D/1D for component c=(t,r,v)
c_EIRmax – maximum energy integral ratio 2D/1D for component c=(t,r,v)
c_T_EIRmax – period of this maximum value
c_lg_EIRmax – maximum energy integral logarithm ratio 2D/1D for component c=(t,r,v)
c_T_lg_EIRm – period of this maximum value
c_RSRmax_A – maximum response spectra for acceleration ratio 2D/1D for component c=(t,r,v)
c_T_RSRm_A – period of this maximum value
c_RSRmax_V – maximum response spectra for for velocity ratio 2D/1D for component c=(t,r,v)
c_T_RSRm_V – period of this maximum value
c_RSRmax_D – maximum response spectra for displacement ratio 2D/1D for component c=(t,r,v)
c_T_RSRm_D – period of this maximum value
c_RSRmax_PSV – maximum response spectra for pseudovelocity ratio 2D/1D for component c=(t,r,v)
c_T_RSRm_PSV – period of this maximum value

In all this cases the code ***"c"*** means that there is a different file for each component.

Annex 6. Ratios 2D/1D for H/V parameters

RsVmaxSA – maximum ratio 2D/1D of the quotient R/V for acceleration
f_RsVmaxSA – frequency at which this maximum occurs
TsVmaxSA – maximum ratio 2D/1D of the quotient T/V for acceleration
f_TsVmaxSA – frequency at which this maximum occurs
HsVmaxSA – maximum ratio 2D/1D of the quotient H/V for acceleration
f_HsVmaSA – frequency at which this maximum occurs
RsVmaxSV – maximum ratio 2D/1D of the quotient R/V for velocity
f_RsVmaSV – frequency at which this maximum occurs
TsVmaxSV – maximum ratio 2D/1D of the quotient T/V for velocity
f_TsVmaSV – frequency at which this maximum occurs
HsVmaxSV – maximum ratio 2D/1D of the quotient R/V for velocity
f_HsVmaSV – frequency at which this maximum occurs
RsVmaxSD – maximum ratio 2D/1D of the quotient R/V for displacement
f_RsVmaSD – frequency at which this maximum occurs
TsVmaxSD – maximum ratio 2D/1D of the quotient T/V for displacement
f_TsVmaSD – frequency at which this maximum occurs
HsVmaxSD – maximum ratio 2D/1D of the quotient H/V for displacement
f_HsVmaSD – frequency at which this maximum occurs
RsVmaxPV – maximum ratio 2D/1D of the quotient R/V for pseudovelocity
f_RsVmaPV – frequency at which this maximum occurs
TsVmaxPV – maximum ratio 2D/1D of the quotient T/V for pseudovelocity
f_TsVmaPV – frequency at which this maximum occurs
HsVmaxPV – maximum ratio 2D/1D of the quotient H/V for pseudovelocity
f_HsVmaPV – frequency at which this maximum occurs

T/V means transverse over vertical components, ***R/V*** means radial over vertical components, while ***H/V*** means horizontal over vertical components.