

# CODAPACK3D. A new version of Compositional Data Package

S. Thió-Henestrosa<sup>1</sup>, O. Gómez<sup>2</sup>, and R. Cepero<sup>2</sup>

<sup>1</sup> Dept. d'Informàtica i Matemàtica Aplicada. Campus Montilivi-Edifici P4, Universitat de Girona, E-17071 Girona. Spain;  
*santiago.thio@udg.edu*

<sup>2</sup> Dept. d'Informàtica i Matemàtica Aplicada. Campus Montilivi-Edifici P4, Universitat de Girona, E-17071 Girona. Spain

## Abstract

The statistical analysis of compositional data should be treated using logratios of parts, which are difficult to use correctly in standard statistical packages. For this reason a freeware package, named CoDaPack was created. This software implements most of the basic statistical methods suitable for compositional data.

In this paper we describe the new version of the package that now is called CoDaPack3D. It is developed in Visual Basic for applications (associated with Excel<sup>®</sup>), Visual Basic and Open GL, and it is oriented towards users with a minimum knowledge of computers with the aim at being simple and easy to use.

This new version includes new graphical output in 2D and 3D. These outputs could be zoomed and, in 3D, rotated. Also a customization menu is included and outputs could be saved in jpeg format. Also this new version includes an interactive help and all dialog windows have been improved in order to facilitate its use.

To use CoDaPack one has to access Excel<sup>®</sup> and introduce the data in a standard spreadsheet. These should be organized as a matrix where Excel<sup>®</sup> rows correspond to the observations and columns to the parts. The user executes macros that return numerical or graphical results. There are two kinds of numerical results: new variables and descriptive statistics, and both appear on the same sheet. Graphical output appears in independent windows. In the present version there are 8 menus, with a total of 38 submenus which, after some dialogue, directly call the corresponding macro. The dialogues ask the user to input variables and further parameters needed, as well as where to put these results. The web site <http://ima.udg.es/CoDaPack> contains this freeware package and only Microsoft Excel<sup>®</sup> under Microsoft Windows<sup>®</sup> is required to run the software.

**Kew words:** Compositional data Analysis, Software.

# 1 Introduction

CoDaPack is a software developed by the Girona group of Compositional Data Analysis (Thió et al., 2003, 2005 and 2006), following the new methodological approach for the statistical analysis of compositional data introduced in the eighties by John Aitchison (1986) and extended by other authors like: Martín-Fernández and others (2000), Barceló-Vidal and others (2001), Pawlowsky-Glahn and Egozcue (2001) and Egozcue and others (2003).

At this moment CoDaPack3D includes some basic statistical methods: transformations, operations within the simplex, zero replacement techniques, ternary diagrams, descriptive statistics, reduction-of-dimension techniques, and some multivariate analysis methods. It is developed in Visual Basic for applications (associated with Excel<sup>®</sup>), Visual Basic and Open GL and it is oriented towards users with a minimum knowledge of computers. It aims to be simple and easy to use.

To use CoDaPack3D one has to execute Excel<sup>®</sup> and open the CoDaPack3D.xls in order to work in a standard spreadsheet of this file. The data inside CoDaPack3D should be organized as a matrix where rows correspond to the observations and columns to the parts. The user executes macros that return numerical or graphical results. There are two kinds of numerical results: new variables and descriptive statistics, and both appear on the same sheet. Graphical output appears in independent windows and uses the OpenGL facilities. The latest version of this freeware package can be found on the web site <http://ima.udg.es/CoDaPack>. Microsoft Excel<sup>®</sup> under Microsoft Windows<sup>®</sup> is required to run it.

## 2 New features of CoDaPack

This new version includes new graphical output in 2D and 3D. Also this new version includes an interactive help and all dialog windows have been improved in order to facilitate its use.

### 2.1 New graphical output

The main change of new CoDaPack is the new graphical output, for this reason the name of the package has changed and now is called CoDaPack3D as it performs 3 dimensional graphs. For this reason two more main menus appear on this version *2D Graphs* and *3D graphs* that contains all this new routines. Also the old routines with old graphical output are maintained inside a main menu now called *Graphs (Old Version)*.

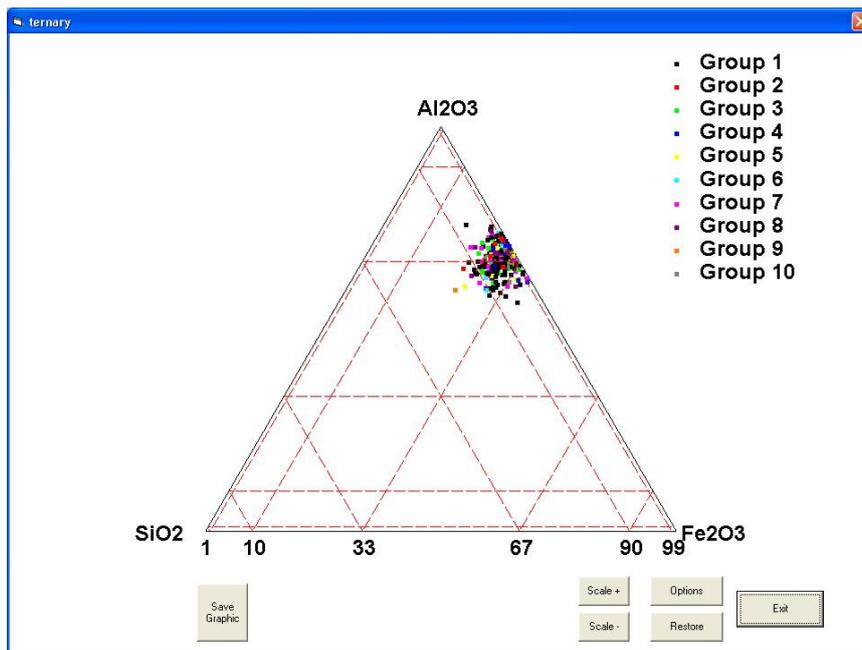


Figure 1: 2D Ternary Diagram working window.

These graphical outputs were performed with OpenGL facilities associated to Visual Basic. When a routine of graphs menu is called, a new working window (Fig. 1) appears with the desired graph. This window is independent of Excel and it is possible to interact with the graph by means of scaling them and, in 3D, rotating them. Also it is possible to change the appearance of label with the *Options* button (Fig. 2).

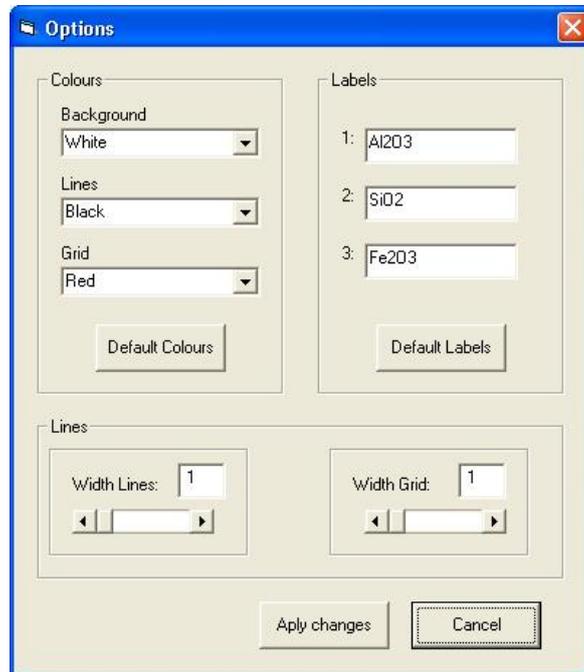


Figure 2: Form of the Options *button* of 2D Ternary Diagram working window.

When the user is working with the graph, every time he wants to restart working on the initial position the user could click on the *Restore* button, and once the graph is correct the user can generate a jpeg file by means of the *Save Graphic* button.

### 2.1.1 2D graphs

Figure 1 shows the working window of a Ternary Diagram output. With the buttons *Scale+* and *Scale-* user can zoom in or zoom out the ternary diagram. Figure 3 shows the ternary diagram after a zoom, and Figure 4 shows a saved jpeg graph without the work buttons.

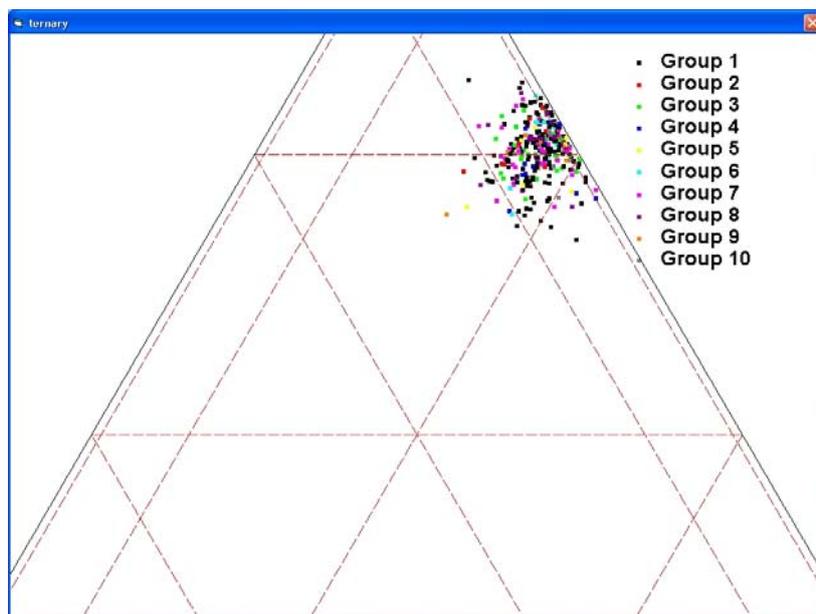


Figure 3: 2D Ternary Diagram after a zoom.

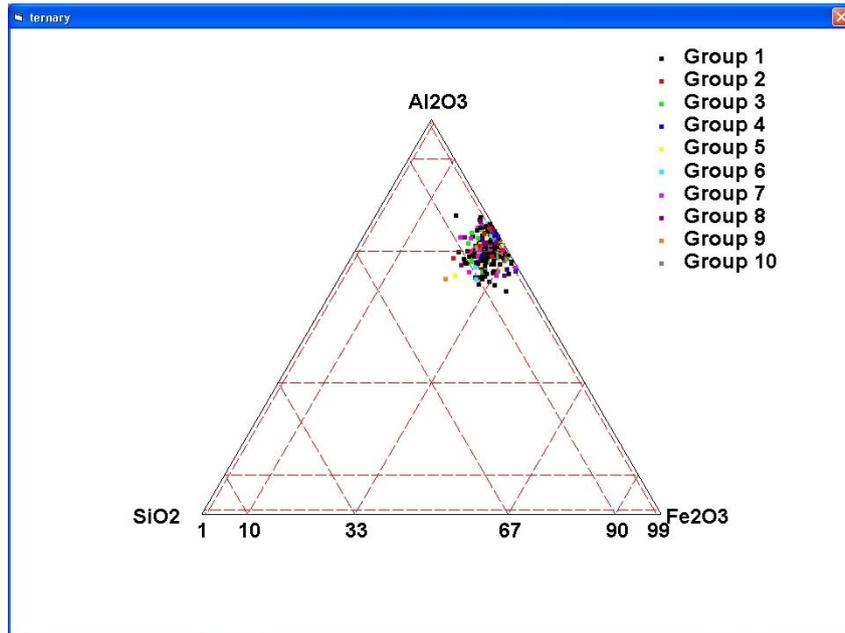


Figure 4: Final jpeg graph of a 2D Ternary Diagram.

### 2.1.2 3D graphs

Some of the graphs of old CoDaPack have been adapted to 3 dimensional ones by adding a new part: Ternary diagram, ALR Plot and CLR Plot. Also, the Biplot could be displayed in 3 dimensions.

Figure 5 shows the working window of a 3D Ternary diagram (formerly a Quaternary diagram) with the zoom buttons (*scale +* and *scale -*) and the rotation buttons (*Rot X +*, *Rot X -*, *Rot Y +*, *Rot Y -*, *Rot Z +* and *Rot Z -*) which right or left rotates on all dimensions. Figure 6A and 6B shows respectively the result of performing a zoom and a rotation.

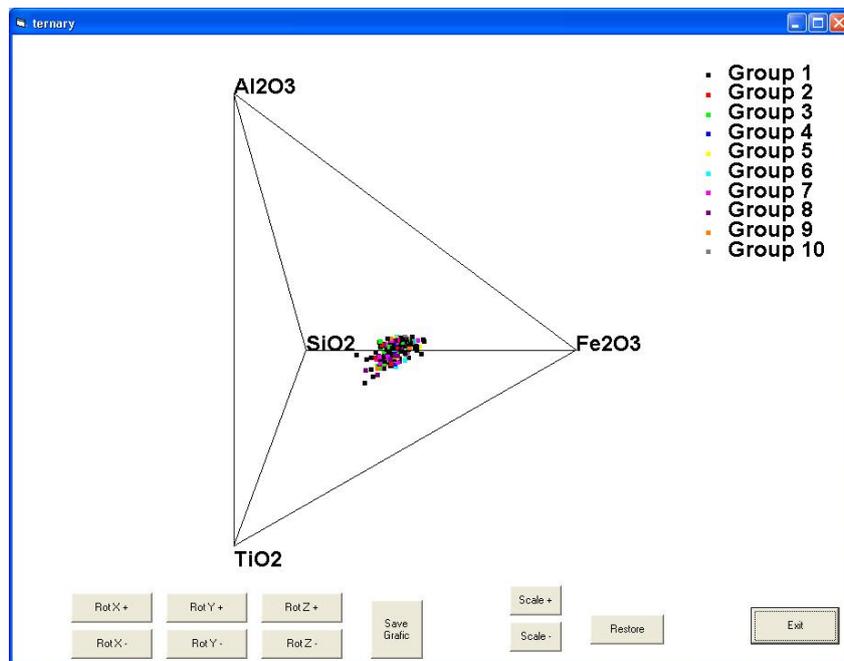


Figure 5: 3D Ternary Diagram working window.

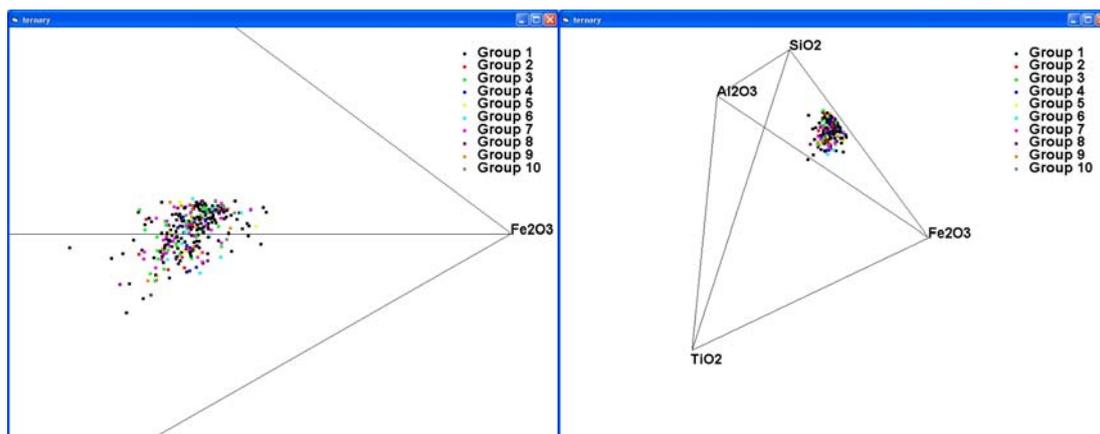


Figure 6: 3D Ternary Diagram: (A) After a zoom; (B) After a rotation.

## 2.2 Other new features

Apart of graphs routines the ILR transformation routine is the only one that has had important changes as it has been divided into two routines, a modified *Raw-ILR* that only performs the direct ILR transformation and *ILR-Raw* that performs the inverse of the ILR. Also in order to clarify its meaning, the names of *Raw-ALR* and *Raw-CLR* are modified to *Raw-ALR-Raw* and *Raw-CLR-Raw* as they performs the transformation in both directions, from simplex to real space and vice versa.

The new *Raw-ILR* routine allows the user to transform the data from the simplex (raw data) to real space (ilr coordinates) with the isometric logratio transformation *ilr* defined by a sequential binary partition.

As in the Balance-Dendrogram routine (Thio et al., 2008), there are three ways of definition of the sequential binary partition: by reading it from the spreadsheet, defined by the user in a special menu and using the default partition.

The *ILR-Raw* routine performs a transformation of the data from real space to the simplex applying the inverse isometric logratio transformation ( $ilr^{-1}$ ) defined by a given sequential binary partition read from the spreadsheet.

## 2.3 New interface

Another significant change of this new version of CoDaPack is the new interface between the program and the users by means of the menus.

### 2.3.1 Easily way to select the parts.

As in the old version of CoDaPack, every time that a menu is selected in CoDaPack3D, a new form appears – standard for all routines, asking which columns to select of the active sheet and where to put the results (Fig. 7). Its left side contains the *Select Columns* structure, the middle the *Inputs* structure and, in some routines, the *Store In (Initial Column)* box and on the right there are always the buttons *Help*, *OK*, *Cancel* and *Clear* and in some routines also other buttons. Between the left and the middle part there are two arrows to pass information between them.

As in old CoDaPack the first row of every column is reserved to labels, when the first form is opened the *Select Columns* list contains the labels row of each column of the Excel sheet and the standard letters that identifies the columns of any Excel sheet beside.

On old versions of CoDaPack to select a part, the user should mark a row of the *Select Columns* list with a single click and then click on the arrow. After that the name of the selected column appears on the middle, inside *Input* structure. The user should repeat this operation in order to select all the parts involved in the routine. Also, to unselect a part of *Inputs* structure the user has to mark this part inside *Input* structure and click the arrow which now indicates the opposite direction.

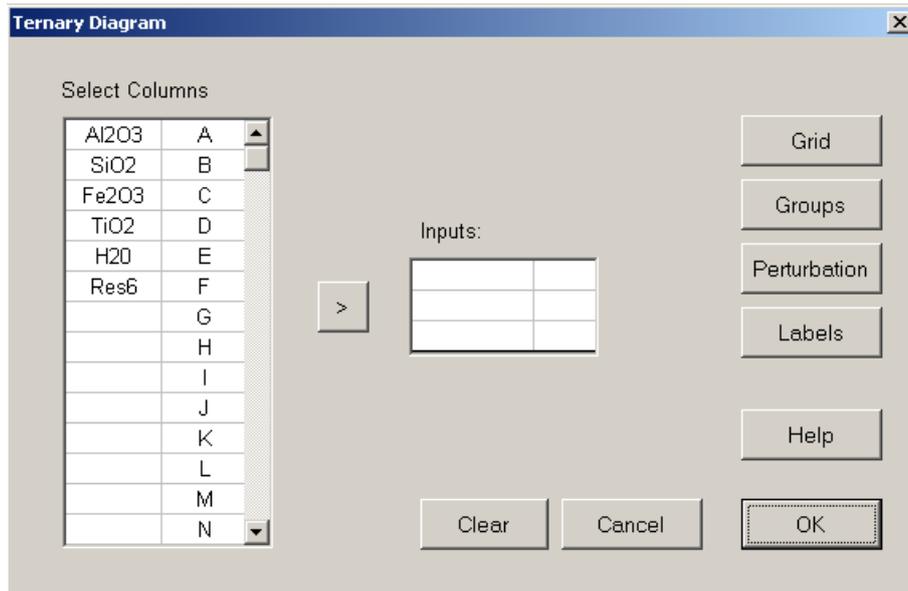


Figure 7: New form of 2D Ternary Diagram menu.

On this new version, CoDaPack3D also allows to select or unselect the parts easily:

- 1) Just double clicking the label or the excel column, the part is selected or unselected depending on where the double click is done (on *Select Columns* or in *Inputs*).
- 2) Performing a multiple selection using the capital letter key and then move it using the arrow or typing the "Return" key to Select from *Select Columns* or the "Del" key to unselect from *inputs*.

Also when an user clicks over a label or an Excel column letter, they appear in blue colour in order to show that this part is selected.

### 2.3.2 New buttons.

The forms of some menus of the old version of CoDaPack contained a single button called, *Options*, to include all optional features of the routine, such as Groups, Labels, etc.

This new version splits the old *Options* button into different ones that allows to the user to see directly which the options of the menu are, and the user could access directly to them.

As on the old version of CoDaPack, the program remembers the options used last time the routine was executed and the *Clear* button resets to the default values.

## 2.4 Help

The last important change performed in CoDaPack is the inclusion of a Help. This help could be called by means of a separate menu or clicking the *Help* button inside every CoDaPack3D routine.

The help is divided into two parts. The first part is a general explanation about general features of CoDapack3D. And the second part is a detailed explanation of all CoDaPack3D menus and contains, for every routine, four main fields:

- 1) Features: This field simply describes the purpose of the routine.
- 2) Buttons and Structure: Explains how to manage with the routine and explores all the optional buttons specifying the default values.
- 3) Examples: This field contains examples of the output obtained by the routine.
- 4) See Also: Links to other related routines of CoDaPack3D.

### 3 Future features of CoDaPack3D

CoDaPack3D is software that is still growing. In the future new routines will be programmed by the group of Girona or with collaboration with other groups. At this moment the work around CoDaPack goes in three main directions:

- 1) **New Numerical output window:** In a few months we would start developing a new independent output window in order to give more statistical outputs and to give results with more decimal precision.
- 2) **New statistic capabilities:** The Girona group plans to work, sometimes in collaboration with other groups, on the programming of new statistic routines.
- 3) **Software maintenance:** In this field we are working (or we want to) on several directions in order to improve CoDaPack3D: a) repairing the bugs reported by users, b) introducing comments into the software in English language in order to facilitate the programming of non catalan developers, c) updating the users manual and d) updating the routines to Excel<sup>®</sup> 2007 edition.

### Acknowledgements and appendices

This work has received financial support from the Dirección General de Investigación of the Spanish Ministry for Science and Technology through the projects BFM2006-03040 and MTM2006-03040.

### References

- Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. Monographs on Statistics and Applied Probability. Chapman & Hall Ltd.
- Barceló-Vidal, C., Martín-Fernández, J.A. and Pawlowsky-Glahn, V. (2001). Mathematical foundations of compositional data analysis. In G. Ross (Ed.), *Proceedings of IAMG'01 --- The sixth annual conference of the International Association for Mathematical Geology*. Volume CD-, 20 pp. electronic publication.
- Egozcue, J.J., Pawlowsky-Glahn, V., Mateu Figueras, G. And Barceló-Vidal, C. (2003). Isometric Logratio Transformations for Compositional Data Analysis. *Mathematical Geology* 35(3), 279-300.
- Pawlowsky-Glahn, V. and Egozcue, J.J. (2001). Geometric approach to statistical analysis on the simplex. *Stochastic Environmental Research and Risk Assessment* 15, p. 384-398.
- Thió-Henestrosa, S., Barceló-Vidal, C., Martín-Fernández, J.A., and Pawlowsky-Glahn, V. (2003). CoDaPack. An Excel and Visual Basic Based Software of Compositional Data Analysis. Current Version and Discussion for Upcoming Versions. In S. Thió-Henestrosa and J.A. Martín-Fernández, (Eds.). *Proceedings of CODAWORK'03, The First Compositional Data Analysis Workshop: Girona, Spain*, 8p. (CD, electronic publication).
- Thió-Henestrosa, S. and Martín-Fernández, J.A. (2005). Dealing with Compositional Data: The Freeware CoDaPack. *Mathematical Geology* 37(7), 777-797.
- Thió-Henestrosa, S. and Martín-Fernández, J.A. (2006). Detailed guide to CoDaPack: a freeware compositional software. In A. Buccianti, G. Mateu-Figueras and V. Pawlowsky-Glahn (Eds.), *Compositional Data Analysis in the Geosciences: from Theory to Practice*. Geological Society, London, Special Publications, 264, 101-118.
- Thió-Henestrosa, S., Egozcue, J.J., Pawlowsky-Glahn, V., Kovács, L.Ó. and Kovács, G.P. (2008). Balance-dendrogram. A new routine of CoDaPack. *Computer and Geosciences*. In press. Electronic version: doi:10.1016/j.physletb.2003.10.071.