

Research Group in Statistics and Data Analysis ([GRCT35-UdG](#)) of the Computer Science and Applied Mathematics Department at the University of Girona.

Members:

Name	e-mail
Dra. Vera Pawlowsky Glahn	vera.pawlowsky@udg.edu
Dr. Carles Barceló Vidal	carles.barcelo@udg.edu
Dr. Santiago Thió Fdez. de Henestrosa	santiago.thio@udg.edu
Dr. Josep Antoni Martín Fernández	josepantoni.martin@udg.edu
Dr. Josep Daunis i Estadella	josep.daunis@udg.edu
Dra. Glòria Mateu Figueras	gloria.mateu@udg.edu

The central idea that unifies the Group is the statistical analysis of **compositional data**. In many scientific disciplines we are used to work with vectors whose components represent relative parts of a whole. Our research addresses the mathematical foundations of the statistical analysis of this kind of data, and also some of the wide-range applications. From the vast previous work of Professor J. Aitchison, built up from a statistical point of view since the early eighties, we have seen that rigorous mathematical foundations are needed to properly understand and apply these techniques. We base them on the definition of a peculiar geometric structure for the **Simplex** (the support space of compositional data), which is in fact a Euclidean space. Afterwards, traditional statistical techniques (cluster, discriminant, factor analysis, regression models, etc.) are adapted to this structure: we are covering both their real-case application and theoretical development, from measure theory or algebraic geometry to differential calculus in the simplex.

Furthermore, we are developing our own statistical package, the [CoDaPack](#) (http://ima.udg.edu/Recerca/EIO/inici_eng.html). This software is oriented to users coming from the applied sciences, with no extensive background in using various computer packages. For this reason it is based on Excel. The features of this new software are very wide: 1) Transformations between the real space to the simplex or viceversa. 2) Operations inside the simplex like centering, subcomposition or rounded zero replacement. 3) 2-D graphical outputs like ternary diagrams or biplots.

The Group has organized three Workshops on Compositional Data Analysis (CoDaWork: 2003, 2005, 2008: <http://ima.udg.edu/Activitats/CoDaWork08/>). The primary goal of the workshop is to identify important potential lines of future research and gain insight as to how they might be tackled. With this goal in mind, we intend to bring together specialist researchers, postgraduate students, data analysts as well as those with a general

interest in the field, prepared to summarize and share crucial contributions and recent developments and to participate as members of a brainstorming think tank.

You can find more information at CoDaWeb [<http://compositionaldata.com/>], the compositional data web site. This web site serves as a forum for the exchange of information and of ideas of all those interested in compositional data analysis. This is a hot topic of research due to two main factors: (1) the long time it took to find a solution to the problem of how to perform a proper statistical analysis of this type of data, i.e. to solve the "spurious correlation problem", as it was named by Karl Pearson back in 1897, or the "closure problem" as it was called by Felix Chayes in the 1960's; and (2) its broad impact in fields as diverse as Geology, Petrology, Chemistry, Medicine, Biology, Economics, or Archaeometry, among others.

Most important papers published in the last three years:

- 1) Thió-Henestrosa, S., Egozcue, J.J., Pawlowsky-Glahn, V. Kovacs, L.O. and Kovacs, G.P. (2007). Balance-dendrogram. A new routine of CoDaPack. *Computers & Geosciences*, ISSN 0098-3004, (accepted for publication).
- 2) Egozcue, J. J., Pawlowsky-Glahn, V. and Díaz-Barrero (2008). Otros espacios euclídeos. *Gaceta de la Real Sociedad Matemática Española* (accepted for publication).
- 3) Mateu-Figueras, G. and Pawlowsky-Glahn, V. (2007). The skew-normal distribution on S^D . *Communications in Statistics - Theory and Methods* (Special issue on Simulation and Computation of skew elliptical distributions), ISSN: 1532-415X (electronic) 0361-0926 (paper), 36(9) (in press).
- 4) Mateu-Figueras, G., Puig, P. and Pewsey, A. (2007). Goodness-of-fit tests for the skew-normal distribution with unknown parameters. *Communications in Statistics - Theory and Methods* (Special issue on Simulation and Computation of skew elliptical distributions), ISSN: 1532-415X (electronic) 0361-0926 (paper), 36(9) (in press).
- 5) Tolosana-Delgado, R. and Pawlowsky-Glahn, V. (2007). Kriging Regionalized Positive Variables Revisited: Sample Space and Scale Considerations. *Mathematical Geology* 39: 529–558. DOI
- 6) 10.1007/s11004-007-9107-7.
- 7) Egozcue, J. J., Pawlowsky-Glahn, V., Ortego, M. I., and Tolosana-Delgado, R. (2006). The effect of scale in daily precipitation hazard assessment. *Natural Hazards and Earth System Sciences*, ISSN 1684-9981, 6: 459-470.
- 8) Buccianti, A. and Pawlowsky-Glahn, V. (2006). Statistical evaluation of compositional changes in volcanic gas chemistry: a case study. *Stochastic Environmental Research and Risk Assessment*, ISSN 1436-3259, 21 (1): 25-33.

- 9) Buccianti, A., Mateu-Figueras, G., and Pawlowsky-Glahn, V. (eds.), *Compositional Data Analysis: from Theory to Practice*, The Geological Society, London, UK. (en prensa)
- 10) Buccianti, A., Mateu-Figueras, G. and Pawlowsky-Glahn, V. (en prensa); *Frequency distributions and natural laws in Geochemistry*, In: Buccianti, A., Mateu-Figueras, G., and Pawlowsky-Glahn, V. (eds.), *Compositional Data Analysis: from Theory to Practice*, The Geological Society, London, UK.
- 11) Daunis-i-Estadella, J. and Barceló-Vidal, C. and Buccianti, A. (en prensa); *Exploratory Compositional Data Analysis*, In: Buccianti, A., Mateu-Figueras, G., and Pawlowsky-Glahn, V. (eds.), *Compositional Data Analysis: from Theory to Practice*, The Geological Society, London, UK.
- 12) Egozcue, J.J. and Pawlowsky-Glahn, V. (en prensa); *Simplicial Geometry for Compositional Data*, In: Buccianti, A., Mateu-Figueras, G., and Pawlowsky-Glahn, V. (eds.), *Compositional Data Analysis: from Theory to Practice*, The Geological Society, London, UK.
- 13) Kolb, C., J. A. Martín-Fernández, R. Abart, H. Lammer; *The chemical variability at the surface of Mars, Icarus* (en prensa).
- 14) Martín-Fernández, J.A. and Thió-Henestrosa, S., (en prensa); *Rounded zeros: some practical aspects for compositional data*, In: Buccianti, A., Mateu-Figueras, G., and Pawlowsky-Glahn, V. (eds.), *Compositional Data Analysis: from Theory to Practice*, The Geological Society, London, UK.
- 15) Pawlowsky-Glahn, V. and Egozcue, J.J. (en prensa); *Compositional data and their analysis: an introduction*, In: Buccianti, A., Mateu-Figueras,
- 16) G., and Pawlowsky-Glahn, V. (eds.), *Compositional Data Analysis: from Theory to Practice*, The Geological Society, London, UK.
- 17) Thió-Henestrosa, S., Martín-Fernández, J.A. (en prensa); *Detailed guide of CoDaPack: a freeware compositional software*, In: Buccianti, A., Mateu-Figueras, G., and Pawlowsky-Glahn, V. (eds.), *Compositional Data Analysis: from Theory to Practice*, The Geological Society, London, UK.
- 18) Egozcue, J. J., Díaz-Barrero, J. L. and Pawlowsky-Glahn, V. (2006). *Hilbert space of probability density functions based on Aitchison geometry*. *Acta Mathematica Sinica*, ISSN 1439-8516, 22(4): 1175-1182.
- 19) Buccianti, A. and Pawlowsky-Glahn, V. (2005). *Water Chemistry and Compositional Data Analysis. New perspectives of investigation*. *Mathematical Geology*, 37(7), 703-727.
- 20) Egozcue, J. J. and Pawlowsky-Glahn, V. (2005). *Groups of parts and their balances in compositional data analysis*. *Mathematical Geology*, 37(7), 799-833.
- 21) Martín-Fernández, J. A., Barceló-Vidal, C., Pawlowsky-Glahn, V., Ó.Kovács, L., and Kovács, G. P. (2005). *Subcompositional patterns in*

Cenozoic volcanic rocks of Hungary. *Mathematical Geology*, 37(7): 729-752.

- 22) Mateu-Figueras, G., Pawlowsky-Glahn, V. and Barceló-Vidal, C. (2005). Additive logistic normal distribution on the simplex. *Stochastic Environmental Research and Risk Assessment (SERRA)*, 19(3): 205-214.
- 23) Thió-Henestrosa, S., Martín-Fernández, J.A. (2005). Dealing with compositional data: the freeware CoDaPack. *Mathematical Geology*, 37(7), 773-793.
- 24) Tolosana-Delgado, R., Otero, N., Pawlowsky-Glahn, V. and Soler, A. (2005). Extracting factor subcompositions from hydrochemical compositions. *Mathematical Geology*, 37(7), 683-702.
- 25) Otero, N., Tolosana-Delgado, R., Soler, A., Pawlowsky-Glahn, V. and Canals, A. (2005). Relative vs. absolute statistical analysis of compositions: a comparative study of surface waters of a Mediterranean river. *Water Research* 39(7): 1404--1414.
- 26) Pawlowsky-Glahn, V., Tolosana-Delgado, R. and Egozcue, J. J. (2005). Scale effect in hazard assessment. Application to daily rainfall. *Advances in Geosciences*, SRef-ID: 1680-7359/adgeo/2005-2-25, 2: 117-121.
- 27) Egozcue, J. J., Pawlowsky-Glahn, V. and Ortego, M. I. (2005). Wave-height hazard analysis in Eastern Coast of Spain. Bayesian approach using generalized Pareto distribution. *Advances in Geosciences*, SRef-ID: 1680-7359/adgeo/2005-2-25, 2: 25-30.