

Aitchison, J.; Egozcue, J. J.

Compositional data analysis: Where are we and where should we be heading? (English)

Zbl 1177.86017

Math. Geol. 37, No. 7, 829-850 (2005).

Summary: We take stock of the present position of compositional data analysis, of what has been achieved in the last 20 years, and then make suggestions as to what may be sensible avenues of future research. We take an uncompromisingly applied mathematical view, that the challenge of solving practical problems should motivate our theoretical research; and that any new theory should be thoroughly investigated to see if it may provide answers to previously abandoned practical considerations.

MSC:

86A32 Geostatistics

Cited in **2** Reviews
Cited in **18** Documents

Keywords:

simplex geometry; Hilbert and Euclidean space; subcomposition; regression; sample space; stay-in-the-simplex

Full Text: DOI

References:

- [1] Aitchison, J., 1981, A new approach to null correlations of proportions: Math. Geol., v. 13, no. 2, p. 175–189. · doi:10.1007/BF01031393
- [2] Aitchison, J., 1982, The statistical analysis of compositional data (with discussion): J. R. Stat. Soc., Ser. B (Stat. Methodol.), v. 44, no. 2, p. 139–177. · Zbl 0491.62017
- [3] Aitchison, J., 1983, Principal component analysis of compositional data: Biometrika, v. 70, no. 1, p. 57–65. · Zbl 0515.62057 · doi:10.1093/biomet/70.1.57
- [4] Aitchison, J., 1984, The statistical analysis of geochemical compositions: Math. Geol., v. 16, no. 6, p. 531–564. · doi:10.1007/BF01029316
- [5] Aitchison, J., 1985, A general class of distributions on the simplex: J. R. Stat. Soc., Ser. B (Stat. Methodol.), v. 47, no. 1, p. 136–146. · Zbl 0582.62014
- [6] Aitchison, J., 1986, The statistical analysis of compositional data. Monographs on statistics and applied Probability: Chapman & Hall, London (Reprinted in 2003 with additional material by Blackburn Press), 416 p. · Zbl 0688.62004
- [7] Aitchison, J., 1990, Relative variation diagrams for describing patterns of compositional variability: Math. Geol., v. 22, no. 4, p. 487–511. · doi:10.1007/BF00890330
- [8] Aitchison, J., 1992a, On criteria for measures of compositional difference: Math. Geol., v. 24, no. 4, p. 365–379. · Zbl 0970.86531 · doi:10.1007/BF00891269
- [9] Aitchison, J., 1992b, The triangle in statistics, in Mardia, K., ed., The art of statistical science. A tribute to G. S. Watson: Wiley, New York, p. 89–104.
- [10] Aitchison, J., 1994, Principles of compositional data analysis, in Anderson, T. W., Olkin, I., and Fang, K., eds., Multivariate analysis and its applications: Institute of Mathematical Statistics, Hayward, CA, p. 73–81.
- [11] Aitchison, J., 1997, The one-hour course in compositional data analysis or compositional data analysis is simple, in Pawlowsky-Glahn, V., ed., Proceedings of IAMG'97–The third annual conference of the International Association for Mathematical Geology, Vol. I, II and addendum: International Center for Numerical Methods in Engineering (CIMNE), Barcelona, Spain, p. 3–35.
- [12] Aitchison, J., 1999, Logratios and natural laws in compositional data analysis: Math. Geol., v. 131, no. 5, p. 563–580. · doi:10.1023/A:1007568008032
- [13] Aitchison, J., 2002, Simplicial inference, in Viana, M. A. G., and Richards, D. S. P., eds., Algebraic methods in statistics and probability, v. 287, Contemporary mathematics series: American Mathematical Society, Providence, RI, p. 1–22.
- [14] Aitchison, J., 2003, Compositional data analysis: Where are we and where should we be heading? See Thió-Henestrosa and Martín-Fernández (2003) (electronic publication).
- [15] Aitchison, J., and Bacon-Shone, J., 1999, Convex linear combination of compositions: Biometrika, v. 86, no. 2, p. 351–364. · Zbl 0931.62009 · doi:10.1093/biomet/86.2.351
- [16] Aitchison, J., and Barceló-Vidal, C., 2002, Compositional processes: A statistical search for understanding: See Bayer, Burger,

- and Skala (2002, p. 381–386).
- [17] Aitchison, J., Barceló-Vidal, C., Egozcue, J. J., and Pawłowsky-Glahn, V., 2002, A concise guide for the algebraic–geometric structure of the simplex, the sample space for compositional data analysis. See Bayer, Burger, and Skala (2002, p. 387–392).
 - [18] Aitchison, J., and Greenacre, M., 2002, Biplots for compositional data: *J. R. Stat. Soc., Ser. C (Appl. Stat.)*, v. 51, no. 4, p. 375–392. · [Zbl 1111.62300](#) · [doi:10.1111/1467-9876.00275](#)
 - [19] Aitchison, J., and Kay, J., 2003, Possible solution of some essential zero problems in compositional data analysis. See Thió-Henestrosa and Martín-Fernández (2003) (electronic publication).
 - [20] Aitchison, J., and Lauder, I. J., 1985, Kernel density estimation for compositional data: *J. R. Stat. Soc., Ser. C (Appl. Stat.)*, v. 34, no. 2, p. 129–137. · [Zbl 0585.62069](#)
 - [21] Aitchison, J., Mateu-Figueras, G., and Ng, K. W., 2004, Characterization of distributional forms for compositional data and associated distributional tests: *Math. Geol.*, v. 35, no. 6, p. 667–680. · [Zbl 1031.62051](#) · [doi:10.1023/B:MATG.0000002983.12476.89](#)
 - [22] Aitchison, J., and Ng, K. W., 2003, Compositional hypotheses of subcompositional stability and specific perturbation change and their testing. See Thió-Henestrosa and Martín-Fernández (2003) (electronic publication).
 - [23] Aitchison, J., and Shen, S. M., 1980, Logistic-normal distributions. Some properties and uses: *Biometrika*, v. 67, no. 2, p. 261–272. · [Zbl 0433.62012](#) · [doi:10.2307/2335470](#)
 - [24] Aitchison, J., and Thomas, C. W., 1998, Differential perturbation processes: A tool for the study of compositional processes. See Buccianti, Nardi, and Potenza (1998, p. 499–504).
 - [25] Azzalini, A., and Capitanio, A., 1999, Statistical applications of the multivariate skew-normal distribution: *J. R. Stat. Soc., Ser. B (Stat. Methodol.)* v. 61, no. 3, p. 579–602. · [Zbl 0924.62050](#)
 - [26] Azzalini, A., and Dalla Valle, A., 1996, The multivariate skew-normal distribution: *Biometrika*, v. 83, no. 4, p. 715–726. · [Zbl 0885.62062](#) · [doi:10.1093/biomet/83.4.715](#)
 - [27] Bacon-Shone, J., 1992, Ranking methods for compositional data: *Appl. Stat.*, v. 41, no. 3, p. 533–537. · [Zbl 0825.62387](#) · [doi:10.2307/2348087](#)
 - [28] Bacon-Shone, J., 2003, Modelling structural zeros in compositional data. See Thió-Henestrosa and Martín-Fernández (2003) (electronic publication).
 - [29] Barceló, C., Pawłowsky-Glahn, V., and Grunsky, E., 1996, Some aspects of transformations of compositional data and the identification of outliers: *Math. Geol.*, v. 28, no. 4, p. 501–518. · [doi:10.1007/BF02083658](#)
 - [30] Barceló-Vidal, C., Martín-Fernández, J. A., and Pawłowsky-Glahn, V., 2001, Mathematical foundations of compositional data analysis, in Ross, G., ed., *Proceedings of IAMG’01–The sixth annual conference of the International Association for Mathematical Geology*, CD-ROM, 20 p. · [Zbl 1101.86310](#)
 - [31] Bayer, U., Burger, H., and Skala, W., eds., 2002, *Proceedings of IAMG’02–The eighth annual conference of the International Association for Mathematical Geology*, Terra Nostra, no. 3
 - [32] Billheimer, D., Guttorp, P., and Fagan, W., 1997, Statistical analysis and interpretation of discrete compositional data: Technical report, NRCSE technical report 11: University of Washington, Seattle, Washington, 48 p. · [Zbl 1073.62573](#)
 - [33] Billheimer, D., Guttorp, P., and Fagan, W., 2001, Statistical interpretation of species composition: *J. Am. Stat. Assoc.*, v. 96, no. 456, p. 1205–1214. · [Zbl 1073.62573](#) · [doi:10.1198/016214501753381850](#)
 - [34] Box, G. E. P., and Cox, D. R., 1964, The analysis of transformations: *J. R. Stat. Soc., Ser. B (Stat. Methodol.)*, v. 26, no. 2, p. 211–252. · [Zbl 0156.40104](#)
 - [35] Buccianti, A., Nardi, G., and Potenza, R., eds., 1998, *Proceedings of IAMG’98–The fourth annual conference of the International Association for Mathematical Geology*, Vol. I and II: De Frede Editore, Napoli, 969 p.
 - [36] Buccianti, A., and Pawłowsky-Glahn, V., 2003, Random variables and geochemical processes: A way to describe natural variability: in Ottonello, G., and Serva, L., *Geochemical baselines of Italy*, Chapter 4: Pacini Editore, Genova, Italy, 294 p. · [Zbl 1103.62111](#)
 - [37] Buccianti, A., Pawłowsky-Glahn, V., Barceló-Vidal, C., and Jarauta-Bragulat, E., 1999, Visualization and modeling of natural trends in ternary diagrams: A geochemical case study. See Lippard, Næss, and Sinding-Larsen (1999, p. 139–144).
 - [38] Buccianti, A., Vaselli, O., and Nisi, B., 2003, New insights on river water chemistry by using noncentred simplicial principal component analysis: A case study. See Thió-Henestrosa and Martín-Fernández (2003) (electronic publication).
 - [39] Butler, J. C., 1979, The effects of closure on the moments of a distribution: *Math. Geol.*, v. 11, no. 1, p. 75–84. · [doi:10.1007/BF01043247](#)
 - [40] Chayes, F., 1960, On correlation between variables of constant sum: *J. Geophys. Res.*, v. 65, no. 12, p. 4185–4193. · [doi:10.1029/JZ065i012p04185](#)
 - [41] Daunis-i-Estadella, J., Egozcue, J. J., and Pawłowsky-Glahn, V., 2002, Least squares regression in the simplex. See Bayer, Burger, and Skala (2002, p. 411–416).
 - [42] Egozcue, J. J., and Pawłowsky-Glahn, V., 2005, Groups of parts and their balances in compositional data analysis. *Math. Geol.*, v. 37, no. 7, p. 795–828. · [Zbl 1177.86018](#)
 - [43] Egozcue, J. J., Pawłowsky-Glahn, V., Mateu-Figueras, G., and Barceló-Vidal, C., 2003, Isometric logratio transformations for compositional data analysis: *Math. Geol.*, v. 35, no. 3, p. 279–300. · [Zbl 1302.86024](#) · [doi:10.1023/A:1023818214614](#)
 - [44] Fry, J. M., Fry, T. R. L., and McLaren, K. R., 2000, Compositional data analysis and zeros in micro data: *Appl. Econ.*, v. 32, no. 8, p. 953–959. · [doi:10.1080/000368400322002](#)
 - [45] Gabriel, K. R., 1971, The biplot–graphic display of matrices with application to principal component analysis: *Biometrika*, v. 58, no. 3, p. 453–467. · [Zbl 0228.62034](#) · [doi:10.1093/biomet/58.3.453](#)

- [46] Gabriel, K. R., 1981, Biplot display of multivariate matrices for inspection of data and diagnosis, in Barnett, V., ed., *Interpreting multivariate data*: Wiley, New York, p. 147–173.
- [47] Galton, F., 1879, The geometric mean, in *vital and social statistics*: Proc. R. Soc. Lond., v. 29, p. 365–366.
- [48] Lippard, S. J., Næss, A., and Sinding-Larsen, R., eds., 1999, *Proceedings of IAMG'99—The fifth annual conference of the International Association for Mathematical Geology*, Vol. I and II: Tapir, Trondheim, Norway, 784 p.
- [49] Martín-Fernández, J. A., Barceló-Vidal, C., and Pawlowsky-Glahn, V., 2000, Zero replacement in compositional data sets, in Kiers, H., Rasson, J., Groenen, P., and Shader, M., eds., *Studies in classification, data analysis, and knowledge organization*: Springer-Verlag, Berlin, p. 155–160. · [Zbl 1101.62352](#)
- [50] Martín-Fernández, J. A., Bren, M., Barceló-Vidal, C., and Pawlowsky-Glahn, V., 1999, A measure of difference for compositional data based on measures of divergence. See Lippard, Næss, and Sinding-Larsen (1999, p. 211–216).
- [51] Martín-Fernández, J. A., Paladea-Albadalejo, J., and Gómez-García, J., 2003, Markov chain Monte Carlo method applied to rounding zeros of compositional data: First approach. See Thió-Henestrosa and Martín-Fernández (2003) (electronic publication).
- [52] Mateu-Figueras, G., 2003, *Models de distribució sobre el simplex*: PhD Thesis, Universitat Politècnica de Catalunya, Barcelona, Spain.
- [53] Mateu-Figueras, G., Barceló-Vidal, C., and Pawlowsky-Glahn, V., 1998, Modeling compositional data with multivariate skew-normal distributions. See Buccianti, Nardi, and Potenza (1998, p. 532–537).
- [54] Mateu-Figueras, G., and Pawlowsky-Glahn, V., 2003, Una alternativa a la distribució lognormal. See Saralegui and Ripoll (2003) (electronic publication).
- [55] Mateu-Figueras, G., Pawlowsky-Glahn, V., and Martín-Fernández, J. A., 2002, Normal in $(\mathbb{R})^+$ vs. lognormal in $(\mathbb{R})^+$. See Bayer, Burger, and Skala (2002, p. 305–310).
- [56] McAlister, D., 1879, The law of the geometric mean: Proc. R. Soc. Lond., v. 29, p. 367–376. · [Zbl 11.0163.04](#)
- [57] Mosimann, J. E., 1962, On the compound multinomial distribution, the multivariate $\{\beta\}$ -distribution and correlations among proportions: *Biometrika*, v. 49, nos. 1–2, p. 65–82. · [Zbl 0105.12502](#)
- [58] Pawlowsky-Glahn, V., 2003, Statistical modelling on coordinates. See (Thió-Henestrosa and Martín-Fernández, 2003) (electronic publication).
- [59] Pawlowsky-Glahn, V., and Buccianti, A., 2002, Visualization and modeling of subpopulations of compositional data: Statistical methods illustrated by means of geochemical data from fumarolic fluids: *Int. J. Earth Sci. (Geol. Rundschau)*, v. 91, no. 2, p. 357–368. · [doi:10.1007/s005310100222](#)
- [60] Pawlowsky-Glahn, V., and Egozcue, J. J., 2001, Geometric approach to statistical analysis on the simplex: *Stochastic Environ. Res. Risk Assess. (SERRA)*, v. 15, no. 5, p. 384–398. · [Zbl 0987.62001](#) · [doi:10.1007/s004770100077](#)
- [61] Pawlowsky-Glahn, V., and Egozcue, J. J., 2002, BLU estimators and compositional data: *Math. Geol.*, v. 34, no. 3, p. 259–274. · [Zbl 1031.86007](#) · [doi:10.1023/A:1014890722372](#)
- [62] Pawlowsky-Glahn, V., Egozcue, J. J., and Burger, H., 2003, An alternative model for the statistical analysis of bivariate positive measurements, in Cubitt, J., ed., *Proceedings of IAMG'03—The ninth annual conference of the International Association for Mathematical Geology*, CD-ROM: University of Portsmouth, Portsmouth, UK.
- [63] Pearson, K., 1897, Mathematical contributions to the theory of evolution. On a form of spurious correlation which may arise when indices are used in the measurement of organs: Proc. R. Soc. Lond., v. LX, p. 489–502. · [Zbl 28.0209.02](#)
- [64] Renner, R. M., 1993, The resolution of a compositional data set into mixtures of fixed source components: *J. R. Stat. Soc., Ser. C (Appl. Stat.)*, v. 42, no. 4, p. 615–631. · [Zbl 0825.62555](#)
- [65] Saralegui, J., and Ripoll, E., eds., 2003, *Actas del XXVII Congreso Nacional de la Sociedad de Estadística e Investigación Operativa (SEIO)*, CD-ROM: Sociedad de Estadística e Investigación Operativa, Lleida (Spain).
- [66] Sarmanov, O. V., and Vistelius, A. B., 1959, On the correlation of percentage values: *Dokl. Akad. Nauk. SSSR*, v. 126, p. 22–25. · [Zbl 0104.13301](#)
- [67] Thió-Henestrosa, S., and Martín-Fernández, J. A., eds., 2003, *Compositional Data Analysis Workshop—CoDaWork'03*, Proceedings: Universitat de Girona, CD-ROM, ISBN 84-8458-111-X, available at <http://ima.udg.es/Activitats/CoDaWork03/>.
- [68] Thomas, C. W., and Aitchison, J., 1998, The use of logratios in subcompositional analysis and geochemical discrimination of metamorphosed limestones from the northeast and central scottish highlands. See Buccianti, Nardi, and Potenza (1998, p. 549–554).
- [69] Thomas, C. W., and Aitchison, J., 2003, Exploration of geological variability and possible processes through the use of compositional data analysis: An example using Scottish metamorphosed limestones. See Buccianti, Nardi, and Potenza (1998) (electronic publication).
- [70] Tolosana-Delgado, R., Otero, N., Pawlowsky-Glahn, V., and Soler, A., 2005, Extracting latent factor subcompositions from hydrochemical compositions. *Math. Geol.*, v. 37, no. 7, p. 681–702.
- [71] Tolosana-Delgado, R., Palomera-Román, R., Gimeno-Torrente, D., Pawlowsky-Glahn, V., and Thió-Henestrosa, S., 2002, A first approach to the classification of basalts using trace elements. See Bayer, Burger, and Skala (2002, p. 435–440).
- [72] Tolosana-Delgado, R., Pawlowsky-Glahn, V., and Mateu-Figueras, G., 2003, Krigeado de variables positivas. Un modelo alternativo. See Bayer, Burger, and Skala (2002) (electronic publication).
- [73] von Eynatten, H., Barceló-Vidal, C., and Pawlowsky-Glahn, V., 2003, Modelling compositional change: The example of chemical weathering of granitoid rocks: *Math. Geol.*, v. 35, no. 3, p. 231–251. · [doi:10.1023/A:1023835513705](#)
- [74] von Eynatten, H., Pawlowsky-Glahn, V., and Egozcue, J. J., 2002, Understanding perturbation on the simplex: A simple

method to better visualize and interpret compositional data in ternary diagrams: *Math. Geol.*, v. 34, no. 3, p. 249–257. · [Zbl 1031.86008](#) · [doi:10.1023/A:1014826205533](#)

- [75] Weltje, J. G., 1997, End-member modeling of compositional data: Numerical–statistical algorithms for solving the explicit mixing problem: *Math. Geol.*, v. 29, no. 4, p. 503–549. · [doi:10.1007/BF02775085](#)

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.