

Martín-Fernández, Josep Antoni; Hron, Karel; Templ, Matthias; Filzmoser, Peter; Palarea-Albaladejo, Javier

Model-based replacement of rounded zeros in compositional data: classical and robust approaches. (English) Zbl 1255.62116
Comput. Stat. Data Anal. 56, No. 9, 2688-2704 (2012).

Summary: The log-ratio methodology represents a powerful set of methods and techniques for statistical analysis of compositional data. These techniques may be used for the estimation of rounded zeros or values below the detection limit in cases when the underlying data are compositional in nature. An algorithm based on iterative log-ratio regressions is developed by combining a particular family of isometric log-ratio transformations with censored regression. In the context of classical regression methods, the equivalence of the method based on additive and isometric log-ratio transformations is proved. This equivalence does not hold for robust regression. Based on Monte Carlo methods, simulations are performed to assess the performance of classical and robust methods. To illustrate the method, a case study involving geochemical data is conducted.

Reviewer: [Reviewer \(Berlin\)](#)

MSC:

[62G08](#) Nonparametric regression and quantile regression
[62G35](#) Nonparametric robustness
[65C05](#) Monte Carlo methods
[65G50](#) Roundoff error

Cited in **14** Documents

Keywords:

[balances](#); [EM algorithm](#); [log-ratio transformations](#); [robust regression](#); [values below detection limit](#)

Software:

[R](#); [robustbase](#)

Full Text: [DOI](#)

References:

- [1] Aitchison, J., The statistical analysis of compositional data. monographs on statistics and applied probability, (1986), Chapman and Hall Ltd. London, UK, p. 416 (Reprinted 2003 with additional material by The Blackburn Press)
- [2] Aitchison, J.; Barceló-Vidal, C.; Martín-Fernández, J.A.; Pawłowsky-Glahn, V., Logratio analysis and compositional distance, *Mathematical geology*, 32, 3, 271-275, (2000) · [Zbl 1101.86309](#)
- [3] Amemiya, T., Tobit models: a survey, *Journal of econometrics*, 24, 3-61, (1984) · [Zbl 0539.62121](#)
- [4] Barceló-Vidal, C., Aguilar, L., Martín-Fernández, J.A., 2011. Compositional VARIMA Time Series (Chapter 7). In: Pawłowsky-Glahn and Buccianti (2011), pp. 87-103.
- [5] ()
- [6] Daunis-i-Estadella, J., Martín-Fernández, J.A. (Eds.), 2008. Proceedings of CODAWORK'08, the 3rd Compositional Data Analysis Workshop. Universitat de Girona. ISBN: 84-8458-272-4. <http://ima.udg.es/Activitats/CoDaWork08/>, May 27-30 (CD-ROM).
- [7] Egozcue, J.J., Reply to on the harker variation diagrams; by J.A. cortés, *Mathematical geosciences*, 41, 7, 829-834, (2009) · [Zbl 1178.86018](#)
- [8] Egozcue, J.J.; Pawłowsky-Glahn, V., Groups of parts and their balances in compositional data analysis, *Mathematical geology*, 37, 7, 795-828, (2005) · [Zbl 1177.86018](#)
- [9] Egozcue, J.J.; Pawłowsky-Glahn, V., Simplicial geometry for compositional data, (), 145-160 · [Zbl 1156.86307](#)
- [10] Egozcue, J.J., Pawłowsky-Glahn, V., 2011. Basic concepts and procedures (Chapter 2). In: Pawłowsky-Glahn and Buccianti (2011), pp. 12-28.
- [11] Egozcue, J.J.; Pawłowsky-Glahn, V.; Mateu-Figueras, G.; Barceló-Vidal, C., Isometric logratio transformations for compositional data analysis, *Mathematical geology*, 35, 3, 279-300, (2003) · [Zbl 1302.86024](#)

- [12] Egozcue, J.J., Tolosana-Delgado, R., Ortego, M.I. (Eds.), 2011. Proceedings of CODAWORK'11, the 4th Compositional Data Analysis Workshop. Sant Feliu De Guxols. ISBN: 978-84-87867-76-7 (electronic publication). May 10-13.
- [13] Filzmoser, P.; Hron, K., Outlier detection for compositional data using robust methods, *Mathematical geosciences*, 40, 3, 233-248, (2008) · [Zbl 1135.62040](#)
- [14] Filzmoser, P., Hron, K., 2011. Robust statistical analysis (Chapter 5). In: Pawlowsky-Glahn and Buccianti (2011), pp. 59-72.
- [15] Filzmoser, P.; Hron, K.; Reimann, C., Principal component analysis for compositional data with outliers, *Environmetrics*, 20, 6, 621-632, (2009)
- [16] Fišerová, E.; Hron, K., On interpretation of orthonormal coordinates for compositional data, *Mathematical geosciences*, 43, 4, 455-468, (2011)
- [17] Hron, K.; Templ, M.; Filzmoser, P., Imputation of missing values for compositional data using classical and robust methods, *Computational statistics and data analysis*, 54, 12, 3095-3107, (2010) · [Zbl 1284.62049](#)
- [18] Huber, P.J., *Robust statistics*, (1981), John Wiley New York · [Zbl 0536.62025](#)
- [19] Johnson, R.A.; Wichern, D.W., *Applied multivariate statistical analysis*, (2002), Prentice Hall London
- [20] Little, R.J.A.; Rubin, D.B., *Statistical analysis with missing data*, (1987), Wiley New Jersey
- [21] Maronna, R.; Martin, R.D.; Yohai, V.J., *Robust statistics: theory and methods*, (2006), John Wiley New York, USA, p. 436 · [Zbl 1094.62040](#)
- [22] Martín-Fernández, J.A.; Barceló-Vidal, C.; Pawlowsky-Glahn, V., Dealing with zeros and missing values in compositional data sets using nonparametric imputation, *Mathematical geology*, 35, 3, 253-278, (2003) · [Zbl 1302.86027](#)
- [23] Martín-Fernández, J.A., Palarea-Albaladejo, J., Olea, R.A., 2011. Dealing with zeros (Chapter 4). In: Pawlowsky-Glahn and Buccianti (2011), pp. 47-62.
- [24] Martín-Fernández, J.A.; Thió-Henestrosa, S., Rounded zeros: some practical aspects for compositional data, (), 191-201
- [25] Mateu-Figueras, G., Barceló-Vidal, C. (Eds.), 2005. Proceedings of CODAWORK'05, the 2nd Compositional Data Analysis Workshop. Universitat de Girona. ISBN: 84-8458-222-1. <http://ima.udg.es/Activitats/CoDaWork05/>, October 19-21 (CD-ROM).
- [26] Mateu-Figueras, G.; Pawlowsky-Glahn, V., A critical approach to probability laws in geochemistry, *Mathematical geosciences*, 40, 5, 489-502, (2008) · [Zbl 1153.86338](#)
- [27] Palarea-Albaladejo, J.; Martín-Fernández, J.A., A modified EM algorithm for replacing rounded zeros in compositional data sets, *Computers & geosciences*, 34, 8, 902-917, (2008)
- [28] Palarea-Albaladejo, J.; Martín-Fernández, J.A.; Gómez-García, J., A parametric approach for dealing with compositional rounded zeros, *Mathematical geology*, 39, 7, 625-645, (2007) · [Zbl 1130.86001](#)
- [29] (), 378
- [30] Pearson, K., 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, *Proceedings of the royal society of London*, 60, 489-502, (1897) · [Zbl 28.0209.02](#)
- [31] R Development Core Team, 2008. R: a language and environment for statistical computing. Vienna. <http://www.r-project.org>.
- [32] Reimann, C.; Åyräs, M.; Chekushin, V.A.; Bogatyrev, I.; Boyd, R.; de Caritat, P.; Dutter, R.; Finne, T.E.; Halleraker, J.H.; Jæger, Ø.; Kashulina, G.; Niskavaara, H.; Lehto, O.; Pavlov, V.; Räsänen, M.L.; Strand, T.; Volden, T., *Environmental geochemical atlas of the central barents region*, (1998), NGU-GTK-CKE Special Publication Trondheim, Norway, p. 745
- [33] Reimann, C.; Filzmoser, P.; Garrett, R.G.; Dutter, R., *Statistical data analysis explained: applied environmental statistics with R*, (2008), Wiley Chichester
- [34] Seber, G.A.F., *A matrix handbook for statisticians*, (2008), John Wiley & Sons, Inc. Hoboken, New Jersey, USA, p. 559
- [35] Thió-Henestrosa, S. Martín-Fernández, J.A. (Eds.), 2003. Proceedings of CODAWORK'03, the 1st Compositional Data Analysis Workshop. Universitat de Girona. ISBN: 84-8458-111-X. <http://ima.udg.es/Activitats/CoDaWork03/>, October 15-17 (CD-ROM).
- [36] Tolosana-Delgado, R., van den Boogaart, K.G., Pawlowsky-Glahn, V., 2011. Geostatistics for compositions (Chapter 6). In: Pawlowsky-Glahn and Buccianti (2011), pp. 73-86.

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.