Pawlowsky-Glahn, V.; **Egozcue, J. J.**

Compositional data in geostatistics: a log-ratio based framework to analyze regionalized compositions. (English) Zbl 1451.86011

Math. Geosci. 52, No. 8, 1067-1084 (2020).

Summary: Problems with compositional data, like spurious correlation and negative bias, are well known in the Geosciences. Not so well known is the fact that the same problems appear when dealing with regionalized compositions. Here, these problems are illustrated, and a solution, based on the principle of working in coordinates using orthonormal logratio representations, is presented. This approach offers a tool for standard geostatistical studies. One of the advantages the method has is that it allows the usual inconsistencies with indicator kriging to be overcome through simplicial indicator kriging. A general way of modelling crossvariograms of coordinates, based on the matrix valued variation variogram, is discussed. In summary, the main aspects related to the modelling and analysis of regionalized compositions have had satisfactory solutions found for them. The proposed methodology is illustrated with public data from a survey concerning arsenic contamination in underground water in Bangladesh.

MSC:

86A32 Geostatistics

Keywords:

compositional data; variation variogram; simplicial indikator kriging; multinomial logistic regression; crossvariogram; compositional kriging

Software:

MASS (R); R; zCompositions; compositions

Full Text: DOI

References:

[1] Aitchison, J., The statistical analysis of compositional data (with discussion), J R Stat Soc Ser B (Stat Methodol), 44, 2, 139-177 (1982) · Zbl 0491.62017

[2] Aitchison, J., Principal component analysis of compositional data, Biometrika, 70, 1, 57-65 (1983) · Zbl 0515.62057 · doi:10.1093/biomet/70.1.57

[3] Aitchison J (1986) The statistical analysis of compositional data. monographs on statistics and applied probability. Chapman \& Hall Ltd., London. (Reprinted in 2003 with additional material by The Blackburn Press). 416 p

[4] Aitchison, J.; Greenacre, M., Biplots for compositional data, J R Stat Soc Ser C (Appl Stat), 51, 4, 375-392 (2002) · Zbl 1111.62300 · doi:10.1111/1467-9876.00275

[5] Aitchison, J.; Kay, JW; Lauder, IJ, Statistical concepts and applications in clinical medicine (2005), Boca Raton: Chapman and Hall/CRC, Boca Raton

[6] Barceló-Vidal, C.; Martín-Fernández, J-A, The mathematics of compositional analysis, Aust J Stat, 45, 57-71 (2016)· doi:10.17713/ajs.v45i4.142

[7] Barceló-Vidal C, Martín-Fernández JA, Pawlowsky-Glahn V (2001) Mathematical foundations of compositional data analysis. In: Ross G (eds) Proceedings of IAMG'01—the VII annual conference of the international association for mathematical geology, Cancun (Mex), pp 20 · Zbl 1052.62531

[8] BGS and DPHE (2001a) Arsenic contamination of groundwater in Bangladesh: data. In: Kinniburgh DG, Smedley PL (eds) Technical report, BGS, DPHE/BGS National Hydrochemical Survey. 1 Excel spreadsheet

[9] BGS and DPHE (2001b) Arsenic contamination of groundwater in Bangladesh. In: Kinniburgh DG, Smedley PL (eds) British geological survey technical report WC/00/19, BGS, Department of Public Health Engineering (Bangladesh). British Geological Survey: Keyworth

[10] Boogaart, K.; Tolosana-Delgado, R., Analysing compositional data with R, 258 (2013), Berlin: Springer, Berlin · Zbl 1276.62011 · doi:10.1007/978-3-642-36809-7

[11] Boogaart, KGV; Tolosana-Delgado, R., "Compositions": a unified r package to analyze compositional data, Comput Geosci, 34, 4, 320-338 (2006)· doi:10.1016/j.cageo.2006.11.017

[12] Chayes, F., Ratio correlation, 99 (1971), Chicago: University of Chicago Press, Chicago

[13] Cox, DR, The regression analysis of binary sequences (with discussion), J R Stat Soc B, 20, 2, 215-242 (1958)

© 2021 FIZ Karlsruhe GmbH

[14] Egozcue JJ, Barceló-Vidal C, Martín-Fernández JA, Jarauta-Bragulat E, Díaz-Barrero JL, Mateu-Figueras G (2011) Elements of simplicial linear algebra and geometry. See Pawlowsky-Glahn and Buccianti (2011), pp 141-157

[15] Egozcue, JJ; Pawlowsky-Glahn, V., Groups of parts and their balances in compositional data analysis, Math Geol, 37, 7, 795-828 (2005) · Zbl 1177.86018 · doi:10.1007/s11004-005-7381-9

[16] Egozcue, JJ; Pawlowsky-Glahn, V.; Gloor, GB, Linear association in compositional data analysis, Aust J Stat, 47, 1, 3-31 (2018)· doi:10.17713/ajs.v47i1.689

[17] Egozcue, JJ; Pawlowsky-Glahn, V.; Mateu-Figueras, G.; Barceló-Vidal, C., Isometric logratio transformations for compositional data analysis, Math Geol, 35, 3, 279-300 (2003) · Zbl 1302.86024 · doi:10.1023/A:1023818214614

[18] Erb I, Notredame C (2016) How should we measure proportionality on relative gene expression data?. Theory Biosci 135(1-2):21-36. doi:10.1007/s12064-015-0220-8

[19] Grunsky, E.; de Caritat, P.; Mueller, U., Using surface regolith geochemistry to map the major crustal blocks of the Australian continent, Gondwana Res, 46, 227-239 (2017)· doi:10.1016/j.gr.2017.02.011

[20] Grunsky, E.; Kjarsgaard, B.; Martin-Fernandez, J.; Thiò-Henestrosa, S., Recognizing and validating structural processes in geochemical data, Compositional data analysis, Vol. 187 of Springer proceedings in mathematics and statistics, 85-116 (2016), New York: Springer, New York

[21] Grunsky, E.; Mueller, U.; Corrigan, D., A study of the lake sediment geochemistry of the melville peninsula using multivariate methods: applications for predictive geological mapping, J Geochem Explor, 141, 15-41 (2014)· doi:10.1016/j.gexplo.2013.07.013

[22] Isaaks, EH; Srivastava, RM, An introduction to applied geostatistics, 561 (1989), New York: Oxford University Press, New York

[23] Journel, A., Nonparametric estimation of spatial distributions, Math Geol, 15, 3, 445-468 (1983)· doi:10.1007/BF01031292

[24] Lovell, D.; Pawlowsky-Glahn, V.; Egozcue, JJ; Marguerat, S.; Bähler, J., Proportionality: a valid alternative to correlation for relative data, PLoS Comput Biol, 11, 3, e1004075 (2015)· doi:10.1371/journal.pcbi.1004075

[25] Martín-Fernández, JA, Comments on: Compositional data: the sample space and its structure, by Egozcue and Pawlowsky-Glahn, TEST, 28, 3, 653-657 (2019) · Zbl 1428.62233 · doi:10.1007/s11749-019-00672-4

[26] Martín-Fernández JA, Egozcue JJ, Olea RA, Pawlowsky-Glahn V (2020) Units recovery methods in compositional data analysis. Nat Resour Res. doi:10.1007/s11053-020-09659-7

[27] Mateu-Figueras G, Pawlowsky-Glahn V, Egozcue JJ (2011) The principle of working on coordinates. See Pawlowsky-Glahn and Buccianti (2011), pp 31-42. 378 p

[28] Mateu-Figueras, G.; Pawlowsky-Glahn, V.; Egozcue, JJ, The normal distribution in some constrained sample spaces, SORT Stat Oper Res Trans, 37, 1, 29-56 (2013) · Zbl 1296.60002

[29] Matheron G (1962) Traité de géostatistique appliquée, Vol. 1. Bureau de Recherches Géologiques et Minières, Paris; Mémoires du Bureau de Recherches Geologiques et Minières, Orléans. Editions Technip, Paris (F). 333 p

[30] Matheron G (2019) Matheron's theory of regionalised variables. In: Pawlowsky-Glahn V, Serra J (eds) International association for mathematical geology—studies in mathematical geology. Oxford University Press, Oxford, vol 9. ISBN: 9780198835660

[31] Molayemat, H.; Torab, FM; Pawlowsky-Glahn, V.; Morshedi, H.; Egozcue, JJ, The impact of the compositional nature of data on coal reserve evaluation, a case study in parvadeh IV coal deposit, central Iran, Int J Coal Geol, 188, 94-111 (2018)· doi:10.1016/j.coal.2018.02.003

[32] Mueller, U.; Grunsky, E., Multivariate spatial analysis of lake sediment geochemical data; Melville peninsula, Nunavut, Canada, Appl Geochem, 75, 247-262 (2016)· doi:10.1016/j.apgeochem.2016.02.007

[33] Olea RA (2008) Inference of distributional parameters from compositional samples containing nondetects. In: Proceedings of CoDaWork'08, the 3rd compositional data analysis workshop. 20 p

[34] Olea, RA; Luppens, JA; Egozcue, JJ; Pawlowsky-Glahn, V., Calorific value and compositional ultimate analysis with a case study of a Texas lignite, J Coal Geol, 162, 27-33 (2016)· doi:10.1016/j.coal.2016.05.005

[35] Palarea-Albaladejo, J.; Martín-Fernández, JA, zCompositions—R package for multivariate imputation of left-censored data under a compositional approach, Chemometr Intell Lab Syst, 143, 85-96 (2015)· doi:10.1016/j.chemolab.2015.02.019

[36] Pawlowsky, V., On spurious spatial covariance between variables of constant sum, Sci Terre Sér Inf, 21, 107-113 (1984)

[37] Pawlowsky V (1986) Räumliche Strukturanalyse und Schätzung ortsabhängiger Kompositionen mit Anwendungsbeispielen aus der Geologie. Ph. D. thesis, Fachbereich Geowissenschaften, Freie Universität Berlin, Berlin (D). 170 p

[38] Pawlowsky, V.; Burger, H., Spatial structure analysis of regionalized compositions, Math Geol, 24, 6, 675-691 (1992)· doi:10.1007/BF00894233

[39] Pawlowsky-Glahn, V.; Buccianti, A., Compositional data analysis: theory and applications (2011), New York: Wiley, New York · Zbl 1103.62111

[40] Pawlowsky-Glahn, V.; Egozcue, J., Exploring compositional data with the Coda-Dendrogram, Aust J Stat, 40, 1-2, 103-113 (2011)

[41] Pawlowsky-Glahn, V.; Egozcue, JJ, Geometric approach to statistical analysis on the simplex, Stoch Environ Res Risk Assess (SERRA), 15, 5, 384-398 (2001) · Zbl 0987.62001 · doi:10.1007/s004770100077

[42] Pawlowsky-Glahn, V.; Egozcue, JJ, BLU estimators and compositional data, Math Geol, 34, 3, 259-274 (2002) · Zbl 1031.86007 · doi:10.1023/A:1014890722372

[43] Pawlowsky-Glahn, V.; Egozcue, JJ, Spatial analysis of compositional data: a historical review, J Geochem Explor, 164, 28-32 (2016)· doi:10.1016/j.gexplo.2015.12.010

[44] Pawlowsky-Glahn, V.; Egozcue, JJ; Lovell, D., Tools for compositional data with a total, Stat Model, 15, 2, 175-190 (2015) ·

© 2021 FIZ Karlsruhe GmbH

Zbl 07258984 · doi:10.1177/1471082X14535526

[45] Pawlowsky-Glahn, V.; Egozcue, JJ; Olea, RA; Pardo-Igúzquiza, E., Cokriging of compositional balances including a dimension reduction and retrieval of original units, J S Afr Inst Min Metall, 115, 1, 59-72 (2015)· doi:10.17159/2411-9717/2015/v115n1a7

[46] Pawlowsky-Glahn, V.; Egozcue, JJ; Tolosana-Delgado, R., Modeling and analysis of compositional data. Statistics in practice, 272 (2015), Chichester: Wiley, Chichester

[47] Pawlowsky-Glahn V, Olea RA (2004) In: DeGraffenreid JA (ed) Geostatistical analysis of compositional data. Number 7 in studies in mathematical geology. Oxford University Press, Oxford · Zbl 1105.86004

[48] Talebi, H.; Mueller, U.; Tolosana-Delgado, R.; Grunsky, E.; McKinley, J.; de Caritat, P., Surficial and deep earth material prediction from geochemical compositions, Nat Resour Res, 28, 869-891 (2019)· doi:10.1007/s11053-018-9423-2

[49] Tolosana-Delgado R (2006) Geostatistics for constrained variables: positive data, compositions and probabilities. Application to environmental hazard monitoring. Ph. D. thesis, Universitat de Girona (Spain)

[50] Tolosana-Delgado, R.; Boogaart, KGVD, Joint consistent mapping of high-dimensional geochemical surveys, Math Geosci, 45, 8, 983-1004 (2008) · Zbl 1321.86035 · doi:10.1007/s11004-013-9485-y

[51] Tolosana-Delgado, R.; Egozcue, JJ; Pawlowsky-Glahn, V.; Ortiz, JM; Emery, X., Cokriging of compositions: log-ratios and unbiasedness, Geostatistics Chile, 299-308 (2008), Santiago: Gecamin Ltd., Santiago

[52] Tolosana-Delgado, R.; Mueller, U.; Boogaart, K., Geostatistics for compositional data: an overview, Math Geosci, 51, 4, 485-526 (2019) · Zbl 1414.86001 · doi:10.1007/s11004-018-9769-3

[53] Tolosana-Delgado, R.; Pawlowsky-Glahn, V.; Egozcue, JJ, Indicator kriging without order relation violations, Math Geosci, 40, 3, 327-347 (2008) · Zbl 1158.86005 · doi:10.1007/s11004-008-9146-8

[54] Tolosana-Delgado, R.; Pawlowsky-Glahn, V.; Egozcue, JJ, Simplicial indicator kriging, J China Univ Geosci, 19, 1, 65-71 (2008)· doi:10.1016/S1002-0705(08)60025-4

[55] Venables, WN; Ripley, BD, Modern applied statistics with S (2002), New York: Springer, New York· doi:10.1007/978-0-387-21706-2

[56] Walwoort, DJ; de Gruijter, JJ, Compositional kriging: a spatial interpolation method for compositional data, Math Geol, 33, 8, 951-966 (2001) · Zbl 1010.86016 · doi:10.1023/A:1012250107121

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.