

Multivariate statistical analysis and partitioning of sedimentary geochemical data sets : General principles and specific MATLAB scripts

The Faculty of Oregon State University has made this article openly available.
Please share how this access benefits you. Your story matters.

Citation	Pisias, N. G., R. W. Murray, and R. P. Scudder (2013), Multivariate statistical analysis and partitioning of sedimentary geochemical data sets: General principles and specific MATLAB scripts, <i>Geochemistry Geophysics Geosystems</i> , 14, 4015–4020. doi:10.1002/ggge.20247
DOI	10.1002/ggge.20247
Publisher	American Geophysical Union
Version	Version of Record
Citable Link	http://hdl.handle.net/1957/47928
Terms of Use	http://cdss.library.oregonstate.edu/sa-termsfuse



Multivariate statistical analysis and partitioning of sedimentary geochemical data sets: General principles and specific MATLAB scripts

Nicklas G. Pias

College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, Oregon, USA

Richard W. Murray and Rachel P. Scudder

Department of Earth and Environment, Boston University, Boston, Massachusetts, 02215, USA (rickm@bu.edu)

[1] Multivariate statistical treatments of large data sets in sedimentary geochemical and other fields are rapidly becoming more popular as analytical and computational capabilities expand. Because geochemical data sets present a unique set of conditions (e.g., the closed array), application of generic off-the-shelf applications is not straightforward and can yield misleading results. We present here annotated MATLAB scripts (and specific guidelines for their use) for Q-mode factor analysis, a constrained least squares multiple linear regression technique, and a total inversion protocol, that are based on the well-known approaches taken by Dymond (1981), Leinen and Pias (1984), Kyte et al. (1993), and their predecessors. Although these techniques have been used by investigators for the past decades, their application has been neither consistent nor transparent, as their code has remained in-house or in formats not commonly used by many of today's researchers (e.g., FORTRAN). In addition to providing the annotated scripts and instructions for use, we discuss general principles to be considered when performing multivariate statistical treatments of large geochemical data sets, provide a brief contextual history of each approach, explain their similarities and differences, and include a sample data set for the user to test their own manipulation of the scripts.

Components: 4,422 words.

Keywords: geochemistry; statistics; multivariate statistics.

Index Terms: 1984 Statistical methods: Descriptive: Informatics; 4318 Statistical analysis: Natural Hazards; 1051 Sedimentary geochemistry: Geochemistry.

Received 20 May 2013; **Revised** 26 July 2013; **Accepted** 5 August 2013; **Published** 2 October 2013.

Pias, N. G., R. W. Murray, and R. P. Scudder (2013), Multivariate statistical analysis and partitioning of sedimentary geochemical data sets: General principles and specific MATLAB scripts, *Geochem. Geophys. Geosyst.*, 14, 4015–4020, doi:10.1002/ggge.20247.

1. Introduction

[2] Concentrations and fluxes of chemical species in and to marine sediment historically have been shown to provide important constraints on a variety of geologic, tectonic, and paleoceanographic processes [e.g., *Goldberg and Arrhenius*, 1958;

Chester, 2000; *Burdige*, 2006; *Schulz and Zabel*, 2006; and many others]. In particular, the distributions of major, trace, and rare earth elements in sediment, sedimentary rocks, settling particles, and atmospheric dust can be used to discern spatial and temporal patterns at many different scales and environments, including hydrothermal regions of

mid-ocean ridges [e.g., *Dymond, 1981; Dekov et al., 2010*], pelagic records of biological export production [e.g., *Pedersen, 1983; Paytan et al., 1996; Murray et al., 2000*], dust and volcanic ash [e.g., *Rea, 1994; Scudder et al., 2009; Muhs, 2013*]; and continental margins [*Taylor and McLennan, 1985; Sholkovitz, 1988*]. Overall, it is important to understand the concentration in and fluxes to deep-sea sediment in order to define key aspects of geochemical cycling and to evaluate changes in such cycles through time.

[3] The introduction to sedimentary chemistry of so-called “rapid” analytical techniques for the analysis of major, trace, and rare earth elements gained momentum in the late 1960s [e.g., *Potts, 1987*] and has continued to the present day. Early applications of instrumental neutron activation analysis (INAA), and X-ray fluorescence (XRF), followed by flame- and graphite furnace atomic absorption (AA), led to the modern and widespread use of inductively coupled plasma emission spectrometry and mass spectrometry (ICP-ES and ICP-MS) techniques. Whereas publications 40 years ago commonly focused on a few chemical elements, it is not unusual for research contributions in the new millennium to include data on 30 or more elements, including full suites of rare earth elements, and often complemented by an array of radiogenic- or stable isotopes.

[4] The arrival of the digital age to all fields of geochemistry has stimulated the development of databases such as PetDB, NavDat, SedDB, and others, which can be accessed through central portals (e.g., www.earthchem.org) that have fundamentally changed the means by which both high- and low-T geochemistry is practiced. These databases are very powerful and allow easy compilation and comparison of data between heretofore disparate locations, ages, data sets, and publications, and further lead to the generation of new data visualization techniques [e.g., *Yamagishi et al., 2011*].

[5] The paired increase in analytical capabilities and the development of large databases has led to a corresponding rise in the need to implement a consistent set of statistical treatments to best utilize these large and high quality data sets. The breadth of these growing databases highlights the need for a consistent means by which their extensive data holdings can be accurately and consistently analyzed statistically. Commercially available software packages often offer multivari-

ate modules of, for example, factor analysis and multiple linear regression, but these generic approaches commonly are not optimized for geochemical use (e.g., to deal with the closed array, and that all end members must have positive concentrations). Moreover, the codes to these treatments are often not readily available, either for proprietary reasons or otherwise, and thus the disciplined user has no way of moving beyond the “black box” application approach, causing a certain amount of information to be lost.

2. Factor Analysis, Multiple Linear Regression, and Total Inversion

[6] To this end, we here provide three annotated MATLAB scripts that address Q-mode Factor Analysis, Constrained Least Squares multiple linear regression, and Total Inversion modeling. These statistical treatments are optimized for use in sedimentary geochemistry but can also be used by other geochemical communities. As detailed below, these approaches have been used successfully for the past 20–30 years [e.g., *Leinen and Piasias, 1984; Leinen, 1987; Knoop and Owen, 1991; McMurtry et al., 1991; Zhou and Kyte, 1992; Kyte et al., 1993*], including a relatively recent series of papers used by members of our research group for sediments in a variety of locations, including the equatorial Pacific Ocean, Cariaco Basin, Arctic Ocean, and the northwest Pacific Ocean [*Ziegler and Murray, 2007; Ziegler et al., 2007, 2008; Martinez et al., 2009, 2010; Scudder et al., 2009*]. While other multivariate techniques are also useful for other approaches [e.g., mineralogy, *Andrews and Eberl, 2012*; general similarity analysis, *Borchardt, 1974*; principal components analysis, *Vermeesch, 2013*], for specific identification of geochemical sources and their respective contributions these Q-mode factor analysis, constrained least squares multiple linear regression, and total inversion techniques have proven robust over the years.

[7] As noted by *Leinen and Piasias [1984]*, there are three steps to evaluating the role of deep sea sedimentation in geochemical cycles within a given sample array: (1) determination of the number of different components that are responsible for the data set being studied, (2) identification of the composition of these different sources, and (3) the quantification of the abundances of each of these components in each sample of the data set being studied. These steps collectively are referred

to as “partitioning” [Leinen and Piasias, 1984]. These statistical methods thus look at the structure of multivariate data sets, and attempt to answer important questions such as, “How are different variables correlated?” and, “How are samples from different locations and times related?” By making these determinations of correlations and relationships, we can then learn about the geological and oceanographic processes that produce the data set.

[8] We have several goals with this paper. First, we hope to provide some uniformity to the statistical treatment of large sedimentary geochemical data sets, which will allow for improved interstudy comparison. Second, we aim to provide increased transparency to the application of multivariate statistics, as users will have access to the fundamental codes and equations governing the statistical application. Third, we hope that this contribution perhaps stimulates further improvements in these and other statistical applications, and ideally these further improvements would also continue to be transparent and readily available. Fourth, we hope that we are able to provide a ready statistical platform to assist other geochemical users in working with their increasingly large and high quality data sets.

3. Application of MATLAB Scripts

3.1. Overview

[9] We here provide three MATLAB scripts that can be used to partition a multivariate geochemical data set. These methods have been applied to other multivariate data sets as well, including microfossil data sets of species relative abundances where the number of species may range from 10 to 100 [e.g., Imbrie and Kipp, 1971; Piasias and Mix, 1997]. We provide a general overview of each of the three statistical procedures, including (a) Q-Mode Factor Analysis, (b) Constrained Least Squares multiple linear regression, and (c) Total Inversion. In the suite of Appendices, we provide the main MATLAB script for each, as well as any other scripts necessary to run the main script, and the mathematical underpinnings are amply described in publications listed in the references. The main MATLAB script for each also gives specific user instructions, including screen shots showing which (and how) key data files must be provided, as well as how to interpret the output.

[10] Due to the unique nature of individual data sets (e.g., the number of samples and elements an-

alyzed), we recognize the need to impose constraints on the treatment of multivariate statistics [e.g., Reimann *et al.*, 2002]. We recommend performing a simple study of concentrations, elemental ratios, normative calculations, r^2 matrices, x - y diagrams, and other well-established approaches when incorporating multivariate techniques into an interpretive strategy. Doing so will allow comparison with lithologic descriptions and mineralogy, define the boundaries for mass balance and, in general, familiarize the user with the fundamental underpinnings of each data set. Furthermore, multiple runs of each statistical treatment with slight variations of the input terms are essential to assess the sensitivity of the statistical methods. Reimann *et al.* [2002] provide a superb discussion of these and other issues relevant to the application of multivariate statistical techniques to geochemical data sets. In the proper context, multivariate techniques such as those we discuss here can provide additional powerful information for geochemistry.

3.2. Q-Mode Factor Analysis

[11] The first script performs what is known as Q-mode factor analysis [e.g., Imbrie and van Andel, 1964; Klován and Imbrie, 1971; Imbrie and Kipp, 1971; Miesch, 1976; Leinen and Piasias, 1984]. Factor analysis and Principal Component Analysis (PCA) are two related methods for identifying relationships between variables in large sets of data. Although both are used in geosciences, one important difference between the two methods is that factor analysis is based on the correlation structure of the variables while PCA accounts for the maximum variance of all the variables. Thus, PCA will by definition force all variables into the result while factor analysis generates unique groups (or “factors”), each of which behaves differently from the other factors. Therefore, factor analysis does not require all variables to be found in the most common factors [Reimann *et al.*, 2002; Ortiz, 2011], which is a more useful characteristic for geochemical applications.

[12] There are two main types of factor analysis, R-mode and Q-mode. Q-mode is more appropriate for geochemical (and many other) applications and thus is used here. R-mode factor analysis is variable based, that is, the objective is to simplify a matrix of variables by forming a smaller number of composite variables that are linear combinations of the original variables. R-mode therefore explains the maximum amount of variance in a data set, which is useful for measuring one

variable at multiple temporal points. Q-mode, on the other hand, is sample based, that is, the objective is to simplify a large matrix of variables measured on many samples. This method is particularly useful for cases when there are many variables measured at multiple spatial or temporal points [Miesch, 1976]. Q-mode factor analysis is preferred for applications in geochemical paleoceanography or other geochemical studies [e.g., Reimann *et al.*, 2002] because of the underlying hypothesis of these types of geochemical studies, namely, that the geochemical data set represents a mixture of different and distinct chemical sources. One of its challenges is that factor score matrices can contain negative values, which is unrealistic when dealing with chemical compositions. A “varimax” rotation can eliminate the problem of the negative values by rotating the principal component axis so that the variability of the data set is orthogonal, but at its maximum. By performing this rotation, the end-member axes are rotated to be closer to real sample compositions [e.g., Leinen and Piasias, 1984]. Overall, Q-mode factor analysis, combined with the “varimax” rotations, results in a description of the data set as being samples composed of positive contributions of a number of geometrically orthogonal factors that provide a close approximation of the hypothesized mixtures.

[13] The script provided here includes extensions of the Q-mode factor analysis of Klovan and Miesch [1976] and Klovan [1981]. Q-mode factor analysis is used to first simplify the multivariate data sets by describing it with a smaller set of components (sometimes called end-members) that, when identified and mapped, can be used to look at processes that ultimately produce the observed data. Although Q-mode factor analysis is usually overdetermined because the number of samples is much greater than the number of variables or factors, it helps address the first step in partitioning by providing an estimate of the number of different components that are contained in the observed data set.

[14] With the extensions to the original Q-mode factor analysis of Klovan and Miesch [1976] and Klovan [1981], this technique further provides estimates for the second and third steps of partitioning, namely, estimation of the composition of the end-members and of the abundance of each end-member in each sample. However, because of the constraint in Q-mode factor analysis that the end-members are algebraically orthogonal (their vector dot products are zero), the compositions of

Q-mode end-members commonly contain negative compositions, as described above. Obviously, this is not acceptable for geochemistry. A number of studies have provided strategies to adjust Q-mode end-members to address the problem posed by the orthogonality constraint [e.g., Full *et al.*, 1981; Leinen and Piasias, 1984], and the MATLAB script provided here also includes this in its calculations.

3.3. Constrained Least Squares (CLS) Multiple Linear Regression

[15] The second script partitions samples in terms of a specified set of end-member components. Unlike the Q-mode factor analysis, this script requires the user to specify the number of components contained in each sample and the exact composition of each of these end-members. In practice, the information gained from the Q-mode factor analysis (e.g., determination of the number of end-members that can be used to describe the data set, and the approximate composition of these end members), can be used as a starting point for the constrained least squares (CLS) statistical treatments.

[16] This MATLAB script is based on the approach used by Dymond [1981], who solved the partitioning problem using linear programming techniques [e.g., Hadley, 1962] to estimate the abundance of five end-members in surface sediment from the Nazca Plate in the Southeast Pacific. Because all equations were linear, the model fit was based on minimizing the sum of the absolute values of the model residuals. The MATLAB script presented here utilizes a constrained least squares (CLS) linear model whereby the contribution from each end-member is constrained to be greater than or equal to zero while minimizing the sum of squares of the residuals. Much of this CLS theory can be found in Rencher [2002]. Where the Q-mode factor analysis generates estimates of end member compositions, this script requires that the end members be specified. Such specification can be determined by the user from data gathered from the literature, and/or from the results of the Q-mode factor analysis.

[17] Given the composition of the end members, constrained least squares techniques are used to calculate the abundance of each end member in each sample. In general, the number of elements analyzed in each sample is greater than the number of end members and thus the set of linear equations is underdetermined. Because we wish to solve these equations with the constraint that the

contributions of each end member in each sample is greater than or equal to zero, the equations are solved using a constrained least squares approach. Traditional linear regression techniques, such as many commonly available in commercially available software packages, focus on minimizing the sum of squares of the residuals, and thus are less advisable because the positivity constraint on the contribution is not satisfied.

3.4. Total Inversion

[18] This final script is based on the FORTRAN code of *Zhou and Kyte* [1992] and *Kyte et al.* [1993] and partitions using the “total nonlinear inversion techniques” outlined by *Tarantola and Vallette* [1982]. Unlike the previous CLS scripts, which assumes that the composition of end-members is fixed precisely, the total inversion script partitions the data set while allowing the composition of each end-member to vary slightly to maximize the partitioning fit of any one sample. Such variation in end member concentrations is more geochemically realistic from several perspectives. First, the actual end member may differ slightly from published values (e.g., from a nearby volcano), and, second, there may be slight variation(s) in composition with time. From the complete partitioning run, the program calculates the same set of fit statistics as CLS as well as the mean and standard deviation of the end-member compositions.

4. Description and Use of the Scripts

[19] In the supporting information¹, we provide four Appendices, each organized around a central theme, and each consisting of all the files a user will need to perform the statistical analyses. Appendix 1 consists of three files necessary to perform the Q-mode Factor Analysis, Constrained Least Squares, and Total Inversion statistical methods on the well-known surface chemistry data of *Dymond* [1981]. Three hundred and twenty-seven surface sediment samples from the Nazca Plate of the southeast Pacific Ocean were analyzed for eight elements (Al, Si, Mn, Fe, Cu, Ni, Zn, and Ba). Data are reported in ppm on a carbonate-free basis [*Dymond*, 1981] and are in Appendix 1a. These sediment samples can be described as a combination of five sediment end-members: terrigenous, biogenic, hydrothermal, authogenic, and

a dissolution residue. The compositions of these end-members are found in Appendix 1b, and the variances in the *Dymond* [1981] data are in Appendix 1c.

[20] Appendices 2, 3, and 4 are organized around the three multivariate statistical treatments. Within each type of analysis (e.g., Q-mode Factor Analysis, in Appendix 2), there is a key file called “Appendix A2a—Users Guide.” This file is essentially a manual complete with instructions for use of the MATLAB code, screen shots of important steps, and tips for ease of use of the code and data. Parallel files are in Appendices 3 and 4, respectively, for CLS and TI. Additionally, Appendices 2–4 contain separate files for the individual MATLAB scripts needed to perform each statistical treatment. In Appendices 2, 3, and 4, we also provide example output for Factor Analysis, CLS, and TI using the Nazca Plate data set.

[21] All files are formatted for easy copy-and-paste input into MATLAB. The scripts were tested on MATLAB version 7.10.0.499 (R2010a) at Boston University from October 2012 to May 2013.

Acknowledgments

[22] RWM and RPS thank NSF for support through NSF OCE 0958002. We further thank NSF for supporting the SedDB project (NSF OCE 0453958 and OCE 0826197), and our colleagues Kirsten Lehnert, Steve Goldstein, and Annika Johansson, all of Lamont Doherty Earth Observatory, for their collaboration during the development of the SedDB database and ensuing discussions regarding development of tools to work with large databases. We thank Joseph Ortiz (Kent State) and Liviu Giosan (Woods Hole Oceanographic Institution) for their constructive reviews, and A. Dunlea (Boston University) for her comments on the earlier manuscript. We thank other recent graduate students from our Boston University research group (C. L. Ziegler, N. Martinez) for their contributions and research that helped incorporate use of these MATLAB scripts.

References

- Andrews, J. T., and D. D. Eberl (2012), Determination of sediment provenance by unmixing the mineralogy of source-area sediments: The “SedUnMix” program, *Mar. Geol.*, 291–294, 24–33.
- Borchardt, G. A. (1974), The SIMAN coefficient for similarity analysis, *Classif. Soc. Bull.*, 3, 2–8.
- Burdige, D. J. (2006), *Geochemistry of Marine Sediments*, 609 pp., Princeton Univ. Press, Princeton, NJ.
- Chester, R. (2000), *Marine Geochemistry*, 2nd ed., 506 pp., Blackwell Sci., Oxford, U. K.
- Dekov, V. M., J. Cuadros, G. D. Kamenov, D. Weiss, T. Arnold, C. Basak, and P. Rochette (2010), Metalliferous

¹Additional supporting information may be found in the online version of this article.

- sediments from the H. M. S. Challenger voyage (1872–1876), *Geochim. Cosmochim. Acta*, *74*, 5019–5038.
- Dymond, J. (1981), The geochemistry of Nazca Plate surface sediments: An evaluation of hydrothermal, biogenic, detrital and hydrogenous sources, *Geol. Soc. Am. Mem.*, *154*, 133–174.
- Full, W. E., R. Ehrlich, and J. E. Klován (1981), Extended Q-model—Objective definition of external end members in the analysis of mixtures, *Math. Geol.*, *13*(4), 331–344.
- Goldberg, E. D., and G. O. S. Arrhenius (1958), Chemistry of Pacific pelagic sediments, *Geochim. Cosmochim. Acta*, *13*, 153–212.
- Hadley, G. (1962), *Linear Programming*, 540 pp., Addison-Wesley, Reading, MA, USA.
- Imbrie, J., and N. G. Kipp (1971), A new micropaleontological method for quantitative paleoclimatology: Application to a late quaternary caribbean core, in *Late Cenozoic Glacial Ages*, edited by K. K. Turekian, pp. 77–181, Yale Univ. Press, New Haven, CT, USA.
- Imbrie, J., and T. van Andel (1964), Vector analysis of heavy-mineral data, *Geol. Soc. Am. Bull.* *75*(11), 1131–1156.
- Klován, J. E. (1981), A generalization of extended Q-Mode factor analysis to data matrices with variable row sums, *Mathe. Geol.*, *13*(3), 217–224.
- Klován, J. E. and J. Imbrie, (1971), An algorithm and FORTRAN-IV Program for large-scale Q-Mode factor analysis and calculation of factor scores, *Math. Geol.*, *3*(1), 61–77.
- Klován, J. E., and A. T. Miesch (1976), Extended CABFAC and QMODEL computer programs for Q-mode factor analysis of compositional data, *Comput. Geosci.*, *1*(3), 161–178.
- Knoop, P. A., and R. M. Owen (1991), Computer software for the analysis of sediment geochemical data in marine minerals exploration, *Mar. Min.*, *10*, 353–368.
- Kyte, F. T., M. Leinen, G. R. Heath, and L. Zhou (1993), Cenozoic sedimentation history of the central North Pacific: Inferences from the elemental geochemistry of Core LL44-GPC3, *Geochim. Cosmochim. Acta*, *57*, 1719–1740.
- Leinen, M. (1987), The origin of paleochemical signatures in North Pacific pelagic clays: Partitioning experiments, *Geochim. Cosmochim. Acta*, *51*, 305–319.
- Leinen, M. and N. G. Pisias (1984), An objective technique for determining end-member compositions and for partitioning sediments according to their sources, *Geochim. Cosmochim. Acta*, *48*, 47–62.
- Martinez, N. C., R. W. Murray, G. R. Dickens, and M. Kolling (2009), Discrimination of sources of terrigenous sediment deposited to the Central Arctic Ocean through the Cenozoic, *Paleoceanography*, *24*, PA1210, doi:10.1029/2007PA001567.
- Martinez, N. C., R. W. Murray, R. C. Thunell, L. C. Peterson, Muller- F. Karger, L. Lorenzoni, Y. Astor, and R. Varela (2010), Local and regional geochemical signatures of surface sediments of the Cariaco Basin and Orinoco Delta, *Venezuela, Geology*, *38*, 159–162.
- McMurtry, G. M., E. H. De Carlo, and K. H. Kim (1991), Accumulation rates, chemical partitioning, and Q-mode factor analysis of metalliferous sediments from the North Fiji Basin, *Mar. Geol.*, *98*, 271–295.
- Miesch, A. T. (1976), Q-mode factor analysis of geochemical and petrologic data matrices with constant row-sums, *U.S. Geol. Surv. Prof. Pap.*, Washington DC, *574-G*, 47 pp.
- Muhs, D. R. (2013), The geologic records of dust in the quaternary, *Aeolian Res.*, *9*, 3–48.
- Murray R. W., C. Knowlton, M. Leinen, A. C. Mix, and C. H. Polisky (2000), Export production and carbonate dissolution in the central equatorial Pacific Ocean over the past 1 Ma, *Paleoceanography*, *15*, 570–592.
- Ortiz, J. D. (2011), Application of Visible/Near Infrared derivative spectroscopy to Arctic paleoceanography, *IOP Conf. Ser.: Earth Environ. Sci.*, *14*, 012011, doi:10.1088/1755-1315/14/1/012011.
- Paytan, A., M. Kastner, and F. P. Chavez (1996), Glacial to interglacial fluctuations in productivity in the equatorial Pacific as indicated by marine barite, *Science*, *274*, 1355–1357.
- Pedersen, T. F. (1983), Increased productivity in the equatorial Pacific during the last glacial maximum, *Geology*, *11*, 16–19.
- Pisias, N. G., and A. C. Mix (1997), Spatial and temporal oceanographic variability of the eastern equatorial Pacific during the Late Pleistocene: Evidence from radiolaria microfossils, *Paleoceanography*, *12*, 381–393.
- Potts, P. J. (1987), *Handbook of Silicate Rock Analysis*, 622 pp., Chapman and Hall, Glasgow, United Kingdom.
- Rea, D. K. (1994), The paleoclimatic record provided by eolian deposition in the deep-sea: The geologic history of wind, *Rev. Geophys.*, *32*, 159–195.
- Reimann, C., P. Filzmoser, and R. G. Garrett (2002), Factor analysis applied to regional geochemical data: Problems and possibilities, *Appl. Geochem.*, *17*(3), 185–206.
- Rencher, A. C. (2002), *Methods of Multivariate Analysis*, 2nd ed., 708 pp., John Wiley, Hoboken, N. J.
- Schulz, H. D., and M. Zabel (2006), *Marine Geochemistry*, 574 pp., Springer, New York, US.
- Scudder, R. P., R. W. Murray, and T. Plank (2009), Dispersed ash in deeply buried sediment from the northwest Pacific Ocean: An example from the Izu-Bonin Arc (ODP Site 1149), *Earth Planet. Sci. Lett.*, *284*, 639–648, doi:10.1016/j.epsl.2009.05.037.
- Sholkovitz, E. R. (1988), Rare earth elements in the sediments of the North Atlantic Ocean, Amazon Delta, and East China Sea: Reinterpretation of terrigenous input patterns to the oceans, *Am. J. Sci.*, *288*, 236–281.
- Tarantola, A., and B. Valette (1982), Generalized nonlinear inverse problem solved using the least squares criterion. *Rev. Geophys.*, *20*, 219–232.
- Taylor S. R., and S. M. McLennan (1985), *The Continental Crust: Its Composition and Evolution*, 312 p., Blackwell Sci, Oxford, UK.
- Vermeesch, P. (2013), Multi-sample comparison of detrital age distributions, *Chem. Geol.*, *341*, 140–146.
- Yamagishi, Y., K. Suzuki, H. Tamura, H. Yanaka, and S. Tsuboi (2011), Visualization of geochemical data for rocks and sediments in Google Earth: Development of a data converter application for geochemical and isotopic data sets in database systems, *Geochem. Geophys. Geosyst.*, *12*, Q03016, doi:10.1029/2010GC003490.
- Zhou, L., and F. T. Kyte (1992), Sedimentation history of the South Pacific pelagic clay province over the last 85 million years inferred from the geochemistry of Deep Sea Drilling Project Hole 596, *Paleoceanography*, *7*, 441–465.
- Ziegler, C. L., and R. W. Murray (2007), Geochemical evolution of the central Pacific Ocean over the past 56 Ma, *Paleoceanography*, *22*, PA2203, doi:10.1029/2006PA001321.
- Ziegler, C. L., R. W. Murray, S. A. Hovan, and D. K. Rea (2007), Resolving eolian, volcanogenic, and authigenic components in pelagic sediment from the Pacific Ocean, *Earth Planet. Sci. Lett.*, *254*, 416–432.
- Ziegler, C. L., R. W. Murray, T. Plank, and S. R. Hemming (2008), Sources of Fe to the equatorial Pacific Ocean from the Holocene to Miocene, *Earth Planet. Sci. Lett.*, *270*, 258–270, doi:10.1016/j.epsl.2008.03.044.