UNIVERSITÀ DI PARMA
ARCHIVIO DELLA RICERCA

University of Parma Research Repository

Comment on “Beyond principal component analysis: Nonparametric and nonlinear approaches for robust analysis of Gafsa basin groundwater” by Kumada and Takefuji (2025)

This is the peer reviewd version of the followng article:

*Original*
Comment on “Beyond principal component analysis: Nonparametric and nonlinear approaches for robust analysis of Gafsa basin groundwater” by Kumada and Takefuji (2025) / Boschetti, T., Rossi, M.. - In: JOURNAL OF GEOCHEMICAL EXPLORATION. - ISSN 0375-6742. - 284:(2026). [10.1016/j.gexplo.2026.108007]

*Availability:*
This version is available at: 11381/3048048 since: 2026-02-16T15:10:59Z

*Publisher:*
Elsevier

*Published*
DOI:10.1016/j.gexplo.2026.108007

*Terms of use:*

Anyone can freely access the full text of works made available as "Open Access". Works made available

*Publisher copyright*

note finali coverpage

(Article begins on next page)

15 November 2026

**Declaration of interests**

☒The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Revised manuscript (unmarked, editable source file)
Click here to view linked References

# Comment on "Beyond principal component analysis: Nonparametric and nonlinear approaches for robust analysis of Gafsa basin groundwater" by Kumada and Takefuji (2025)

Tiziano Boschetti*, Mattia Rossi

Department of Chemistry, Life Sciences and Environmental Sustainability, University of Parma, Parco Area delle Scienze 157/a, 43124 Parma, Italy

*Corresponding author: tiziano.boschetti@unipr.it

We thank the authors for their interest in our work and for fostering constructive discussion on multivariate methods applied to hydrogeochemical datasets from the Gafsa Basin, Tunisia (Kumada and Takefuji, 2025).

In their abstract, Kumada and Takefuji (2025) state that "complementary methodologies—Feature Agglomeration, Independent Component Analysis, and High Variance Gene Selection—create a more comprehensive analytical framework capable of capturing nonlinear relationships, hierarchical structures, and statistically independent variation sources that PCA might overlook." We encourage the authors to substantiate this claim with quantitative evidence, such as figures or tables derived from the Gafsa Basin dataset, which is publicly available as supplementary material in Boschetti et al. (2025).

We further encourage a more detailed discussion of the assumptions and limitations inherent to the proposed methods. For example, Independent Component Analysis (ICA) requires non-Gaussianity and sufficiently large sample sizes to yield stable and interpretable components. Miki et al. (2025) applied ICA to a dataset with 135 samples and 19 variables ($n/p \approx 7$), whereas the Gafsa dataset comprises 33 samples and 19 variables ($n/p \approx 1.7$). Under such conditions, ICA is more appropriately regarded as an exploratory rather than an inferential tool.

We acknowledge that Principal Component Analysis (PCA) has well-known limitations. However, in hydrogeochemical studies characterized by relatively small sample sizes and moderate-to-high dimensionality, alternative probabilistic approaches such as Exploratory Factor Analysis (EFA) are often impractical. EFA relies on iterative parameter estimation procedures (e.g., maximum likelihood), which in low $n/p$ or $N<p$ settings frequently result in non-convergence, unstable solutions, or improper estimates such as Heywood cases (Widaman, 1993). Moreover, EFA requires adequate sampling conditions, commonly assessed through measures such as the Kaiser–Meyer–Olkin index and Bartlett's test, which are rarely satisfied in sparse environmental datasets.

In contrast, PCA provides a mathematically well-defined and computationally stable framework for dimensionality reduction. By applying a deterministic algebraic transformation of the variance–covariance (or correlation) matrix, PCA yields a unique solution without requiring assumptions about latent variable structures or iterative model fitting (Jolliffe and Cadima, 2016; Velicer and Jackson, 1990). For these reasons, PCA is often preferred for feature extraction and exploratory analysis in contexts where computational robustness and reproducibility take precedence over formal inference (e.g., identification of latent constructs). Accordingly, in Boschetti et al. (2025), PCA was intentionally employed for exploratory and descriptive purposes, with component interpretation constrained by independent hydrogeological and geochemical evidence, including the a posteriori use of a categorical variable related to predetermined pollution sources, rather than being treated as inferential or causal.

PCA has a long-standing and well-established tradition in hydrogeochemical research (e.g., Vriend, 1990; Boschetti et al., 2003; Boschetti et al., 2005; Cortecci et al., 2008; Golla, 2018; Barbieri et al., 2021). Notably, the combined use of PCA and hierarchical cluster analysis (HCA) has consistently proven effective in identifying distinct hydrochemical water types from springs and wells, even in highly complex settings, such as hydrothermal systems in arid climates and under conditions of limited sample availability (e.g., Awaleh et al., 2015; Awaleh et al., 2017; Awaleh et al., 2018; Awaleh et al., 2024). Conversely, we are not aware of applications of any High Variance Gene Selection (HVGS) to hydrogeochemical datasets. HVGS was developed primarily for high-dimensional omics data, where the number of variables is orders of magnitude greater than the number of observations. In hydrogeochemistry, sample sizes are typically constrained by the availability of wells or springs, potentially limiting both the applicability and statistical stability of HVGS-based approaches.

In summary, while PCA—like any multivariate technique—has limitations, it remains a valuable exploratory tool when applied with appropriate caution, particularly in low n/p settings. Other approaches, including ICA and Feature Agglomeration, may provide complementary perspectives but are likewise subject to methodological constraints and stability issues when sample sizes are limited. More broadly, we acknowledge that all multivariate approaches are sensitive to sample size and data structure, and that no single method can fully compensate for instability in low n/p datasets. Consistent with Yang and Chang (2015), we regard multivariate variable-reduction techniques as complementary rather than mutually exclusive, with their suitability depending on data structure, sample size, and study objectives.

**Acknowledgements**

We sincerely thank the two anonymous reviewers for their constructive comments and helpful suggestions, which contributed to improving the clarity and quality of this Comment.

**References**


Awaleh, M.O., Hoch, F.B., Boschetti, T., Soubaneh, Y.D., Egueh, N.M., Elmi, S.A., Mohamed, J., Khaireh, M.A. (2015). The geothermal resources of the Republic of Djibouti—II: Geochemical study of the Lake Abhe geothermal field. Journal of Geochemical Exploration, 159, 129-147. http://dx.doi.org/10.1016/j.gexplo.2015.08.011

Awaleh, M.O., Boschetti, T., Soubaneh, Y.D., Baudron, P., Kawalieh, A. D., Dabar, O.A., Ahmed, M.M., Daoud, M.A., Egueh, N.M., Mohamed, J. (2017). Geochemical study of the Sakalol-Harralol geothermal field (Republic of Djibouti): Evidences of a low enthalpy aquifer between Manda-Inakir and Asal rift settings. Journal of Volcanology and Geothermal Research, 331, 26-52. http://dx.doi.org/10.1016/j.jvolgeores.2016.11.008

Awaleh, M. O., Boschetti, T., Soubaneh, Y. D., Kim, Y., Baudron, P., Kawalieh, A. D., Ahmed, M.M., Daoud, M.A., Dabar, O.A., Kadieh, I.H., Adiyaman, Ö., Elmi, S.A., I.H., Chirdon, M.A. (2018). Geochemical, multi-isotopic studies and geothermal potential evaluation of the complex Djibouti volcanic aquifer (republic of Djibouti). Applied Geochemistry, 97, 301-321. https://doi.org/10.1016/j.apgeochem.2018.07.019

Awaleh, M.O., Boschetti, T., Ahmed, M.M., Dabar, O.A., Robleh, M.A., Waberi, M.M., Ibrahim, N.H., Dirieh, E.S. (2024). Spatial distribution, geochemical processes of high-content fluoride and nitrate groundwater, and an associated probabilistic human health risk appraisal in the Republic of Djibouti. Science of the Total Environment, 927, 171968. https://doi.org/10.1016/j.scitotenv.2024.171968

Barbieri, M., Franchini, S., Barberio, M. D., Billi, A., Boschetti, T., Giansante, L., Gori, F., Jónsson, S., Petitta, M., Skelton, A., Stockmann, G. (2021). Changes in groundwater trace element concentrations before seismic and volcanic activities in Iceland during 2010–2018. Science of the Total Environment, 793, 148635. https://doi.org/10.1016/j.scitotenv.2021.148635

Boschetti, T., Hamed, Y., Hadji, R., Barbieri, M., Gentilucci, M., Rossi, M., Khalil, R., Khan, S.D., Ashgar, B., Al-Omran, A., Elaloui, E. (2025). Using principal component analysis to distinguish sources of radioactivity and nitrates contamination in Southern Tunisian groundwater samples. Journal of Geochemical Exploration, 271, 107670. https://doi.org/10.1016/j.gexplo.2025.107670

Boschetti, T., Cortecci, G., Bolognesi, L. (2003). Chemical and isotopic compositions of the shallow groundwater system of Vulcano Island, Aeolian Archipelago, Italy: an update. GeoActa, 2, 1-34.

Boschetti, T., Venturelli, G., Toscani, L., Barbieri, M., Mucchino, C. (2005). The Bagni di Lucca thermal waters (Tuscany, Italy): an example of Ca-SO4 waters with high Na/Cl and low Ca/SO4 ratios. Journal of Hydrology, 307(1-4), 270-293. https://doi.org/10.1016/j.jhydrol.2004.10.015

Cortecci, G., Dinelli, E., Boschetti, T., Arbizzani, P., Pompilio, L., Mussi, M. (2008). The Serchio River catchment, northern Tuscany: Geochemistry of stream waters and sediments, and isotopic composition of dissolved sulfate. Applied Geochemistry, 23(6), 1513-1543. https://doi.org/10.1016/j.apgeochem.2007.12.031

Golla, J. K. (2018). Using Principal Component Analysis to Aid in Visualization and Interpretation of Geothermal Solute Chemistry: An Application to Yellowstone Thermal Waters. Geothermal Resources Council 42.

Jolliffe, I. T., Cadima, J. (2016). Principal component analysis: a review and recent developments. Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences, 374(2065), 20150202. https://doi.org/10.1098/rsta.2015.0202

Kumada, K., Takefuji, Y. (2025). Beyond principal component analysis: Nonparametric and nonlinear approaches for robust analysis of Gafsa basin groundwater. Journal of Geochemical Exploration, 107962. https://doi.org/10.1016/j.gexplo.2025.107962

Miki, Y., Takazawa, E., Ueki, K., Kuwatani, T. (2025). Variable upper mantle geochemical processes constrained from independent component analysis of the Fizh Massif, Northern Oman Ophiolite. Geochemistry, Geophysics, Geosystems, 26(3), e2024GC012134. https://doi.org/10.1029/2024GC012134

Velicer, W. F., Jackson, D. N. (1990). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. Multivariate behavioral research, 25(1), 1-28. https://doi.org/10.1207/s15327906mbr2501_1

Vriend, S. P. (1990). Practical applications of multivariate statistics in exploration geochemistry. Geologica Ultraiectina, 70, 109 pp.

Widaman, K. F. (1993). Common factor analysis versus principal component analysis: Differential bias in representing model parameters? Multivariate Behavioral Research, 28(3), 263-311. https://doi.org/10.1207/s15327906mbr2803_1

Yang, J., Cheng, Q. (2015). A comparative study of independent component analysis with principal component analysis in geological objects identification, Part I: Simulations. Journal of Geochemical Exploration, 149, 127-135. https://doi.org/10.1016/j.gexplo.2014.11.013