the explanation about the statistical methods

Cluster analysis (CA) is also known as the unsupervised pattern recognition method. We explored the spatiotemporal similarity of water quality at 14 monitoring stations by CA [1]. Considering the squared Euclidean distance as the similarity measure, Ward's method was used to perform hierarchical CA (HCA) on the normalized data set [2].

Discriminant analysis (DA), also known as the supervised pattern recognition method, statistically classifies original cases. The main purpose of DA is to extract the important variables reflecting the variation between groups [3]. DA establishes the discriminant function for each group and uses multiple weighted variables to distinguish the characteristics of different groups as follows [4]:

$$f(G_i) = k_i + \sum_{j=1}^{n} w_{ij} p_{ij} \quad (1)$$

where i is the number of groups (G), k_i is the inherent constant of each group, n is the number of parameters used to classify data, and w_{ij} is the weight, which is assigned to the given parameter (p_{ij}). The classification matrix, also known as the confusion matrix or allocation table, was used to evaluate the performance of DA. The rows in the table are the observed categories, the columns are the predicted categories, and the diagonals are the numbers of cases correctly classified.

When constructing the discriminant function, the standard, forward stepwise and backward stepwise modes were performed on the original data set. Regions (space) and seasons (time) were grouped variables, and all the water quality parameters were independent variables [5,6].

Principal component analysis/factor analysis (PCA/FA) is an effective tool to reduce the dimensionality of large data sets without losing useful information. It obtains eigenvalues and eigenvectors from the covariance matrix of the original variables to generate new orthogonal variables called principal components (PCs) and performs varimax rotation on these PCs to generate new variable groups called variance factors (VFs). PCs are linear combinations of observable variables, while VFs can contain nonobservable hypothetical potential variables [7]. VFs further reduce the contribution of secondary variables, simplifying more data structures from PCA [8,2,9]. FA/PCA can explain the physical, chemical, or biological sources of water contamination in PCs.

References

1. Singh, K.P.; Malik, A.; Mohan, D.; Sinha, S. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)—a case study. *Water Res.* **2004**, *38*, 3980-3992. https://doi.org/10.1016/j.watres.2004.06.011.
2. Vadde, K.; Wang, J.; Cao, L.; Yuan, T.; McCarthy, A.; Raju, S. Assessment of Water Quality and Identification of Pollution Risk Locations in Tiaoxi River (Taihu Watershed), China. *Water* **2018**, *10*, 183. https://doi.org/10.3390/w10020183.
3. Xu, Y.; Xie, R.; Wang, Y.; Jian, S. Spatio-temporal variations of water quality in Yuqiao Reservoir Basin, North China. *Front. Env. Sci. Eng.* **2015**, *9*, 649–664. https://doi.org/10.1007/s11783-014-0702-9.

4. Muangthong, S.; Shrestha, S. Assessment of surface water quality using multivariate statistical techniques: case study of the Nampong river and Songkhram river, Thailand. *Environ. Monit. Assess.* **2015**, *187*, 548. https://doi.org/10.1007/s10661-015-4774-1.
5. Zhang, X.; Wang, Q.; Liu, Y.; Wu, J.; Yu, M. Application of multivariate statistical techniques in the assessment of water quality in the Southwest New Territories and Kowloon, Hong Kong. *Environ. Monit. Assess.* **2011**, *173*, 17–27. https://doi.org/10.1007/s10661-010-1366-y.
6. Feng, Z.; Yong, L.; Guo, H. Application of multivariate statistical methods to water quality assessment of the watercourses in northwestern new territories, Hong Kong. *Environ. Monit. Assess.* **2007**, *132*, 1–13. https://doi.org/10.1007/s10661-006-9497-x.
7. Alberto, W.D.; Pilar, D.D.; Valeria, A.M.; Fabiana, P.S.; Cecilia, H.A.; De, B.M. Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquía River Basin (Cordoba-Argentina). *Water Res.* **2001**, *35*, 2881-2894. https://doi.org/10.1016/S0043-1354(00)00592-3.
8. Shrestha, S.; Kazama, F. Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. *Environ. Modell. Softw.* **2007**, *22*, 464–475. https://doi.org/10.1016/j.envsoft.2006.02.001.
9. Abdul-Wahab, S.A.; Bakheit, C.S.; Al-Alawi, S.M. Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations. *Environ. Modell. Softw.* **2005**, *20*, 1263–1271. https://doi.org/10.1016/j.envsoft.2004.09.001.