Supporting information

Identifying key influences on surface water quality in freshwater areas of the Vietnamese Mekong Delta from 2018 to 2020

Carolin Wehrheim[1*], Manfred Lübken[1], Harro Stolpe[2] and Marc Wichern[1]

[1] Institute of Urban Water Management and Environmental Engineering, Ruhr University Bochum, 44801 Bochum, Germany;
[2] Institute of Environment and Ecology, Ruhr University Bochum, Germany, Ruhr University Bochum, 44801 Bochum, Germany;

[*] Correspondence: carolin.wehrheim@rub.de;

Content

Text S1 Principal component analysis

During a dimension reduction, several variables (e.g., water quality parameters) are combined with different weighting into one larger cluster, the “principal component”. Components are identified that explain most of the variances of a system [1]. In this way, water quality parameters can be assigned into groups based on their variance within the data set and subsequently weighted according to their relevance. The generated groups give information about specific parameters that may represent site-specific pollution. To verify the applicability of the principal component analysis, the data set was tested according to the criteria of the Bartlett test of sphericity and the Kaiser-Meyer-Olkin measure of sampling adequacy (MSA). Parameters that caused multicollinearity or negatively affected the MSA by low variation in concentration were removed from the data set, such as T and NO_3-N. The reduced data set included DO, TSS, BOD_5, COD, NH_4-N, NO_2-N, PO_4-P, Fe_{total} and TC. For the scalar transformation of the calculated covariance matrix of the water quality data, the eigenvectors and eigenvalues are calculated, and the latter ranked in descending order. Since the eigenvalue represents the proportion of the total variance of the data set, only components with eigenvalues > 1 were considered as being relevant for the further analysis. This serves to select only those components that significantly represent the variance of the data set. The selected components were further orthogonally rotated by varimax rotation with Kaiser normalization according to Kaiser (1958) to narrow medium-range correlations between the original variables and the created components [2]. This allows a simplified decision of selecting representative variables for each component [3].

Figure S1 Flow diagram

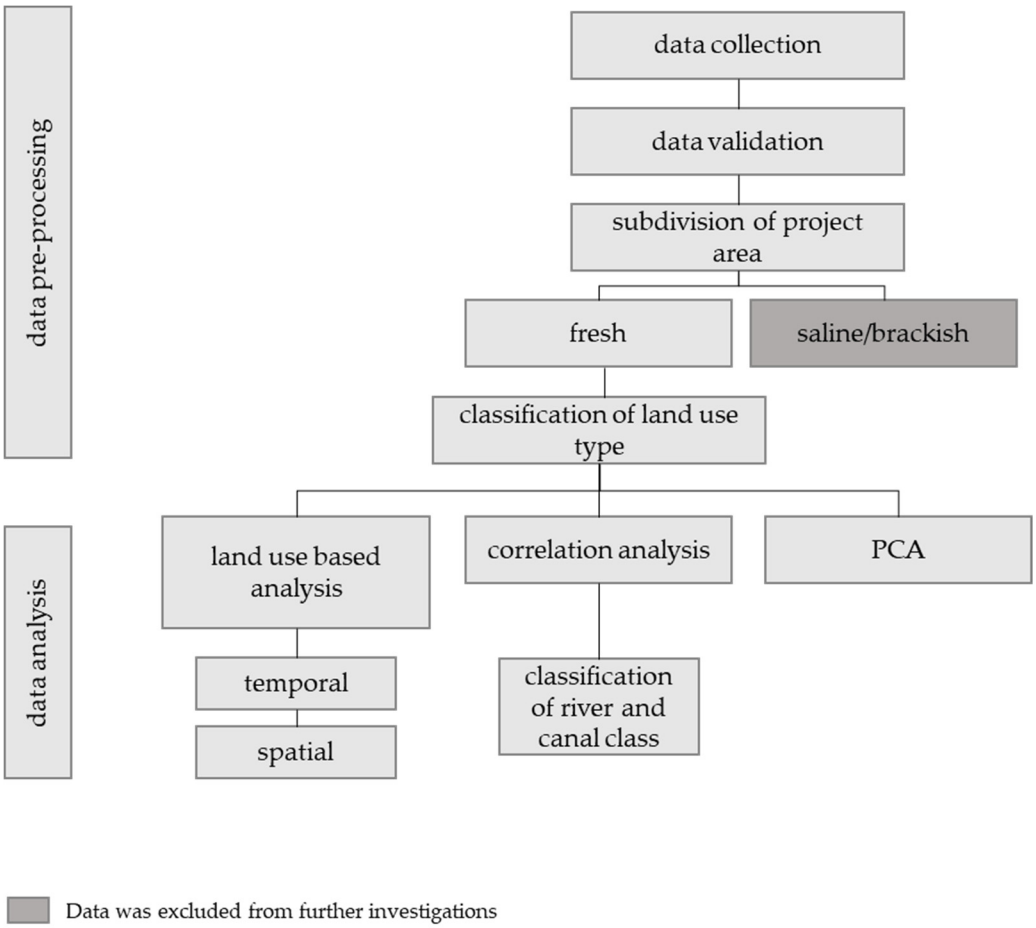

Figure S1: Overview of the steps performed.

References

1. Ouyang, Y. Evaluation of river water quality monitoring stations by principal component analysis. *Water Res.* **2015**, *39*, 2621–2635. https://doi.org/10.1016/j.watres.2005.04.024.
2. Kaiser, H.F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika* **1958**, *23*, 187–200.
3. Chatfield, C.; Collins, A.J. *Introduction to Multivariate Analysis*; Springer: Boston, MA, USA, 1980; ISBN 978-0-412-16030-1.