

Effect of data pre-treatment procedures on principal component analysis: A case study for mangrove surface sediment datasets

Abstract

Principal component analysis (PCA) is capable of handling large sets of data. However, lack of consistent method in data pre-treatment and its importance are the limitations in PCA applications. This study examined pre-treatments methods ($\log(x + 1)$ transformation, outlier removal, and granulometric and geochemical normalization) on dataset of Mengkabong Lagoon, Sabah, mangrove surface sediment at high and low tides. The study revealed that geochemical normalization using AI with outliers removal resulted in a better classification of the mangrove surface sediment than that outliers removal, granulometric normalization using clay and $\log(x + 1)$ transformation. PCA output using geochemical normalization with outliers removal demonstrated associations between environmental variables and tides of mangrove surface sediment, Mengkabong Lagoon, Sabah. The PCA outputs at high and low tides also provided to better interpret information about the sediment and its controlling factors in the intertidal zone. The study showed data pre-treatment method to be a useful procedure to standardize the datasets and reducing the influence of outliers.