Robust principal component analysis in water quality index development

ABSTRACT

Some statistical procedures already available in literature are employed in developing the water quality index, WQI. The nature of complexity and interdependency that occur in physical and chemical processes of water could be easier explained if statistical approaches were applied to water quality indexing. The most popular statistical method used in developing WQI is the principal component analysis (PCA). In literature, the WQI development based on the classical PCA mostly used water quality data that have been transformed and normalized. Outliers may be considered in or eliminated from the analysis. However, the classical mean and sample covariance matrix used in classical PCA methodology is not reliable if the outliers exist in the data. Since the presence of outliers may affect the computation of the principal component, robust principal component analysis, RPCA should be used. Focusing in Langat River, the RPCA-WQI was introduced for the first time in this study to re-calculate the DOE-WQI. Results show that the RPCA-WQI is capable to capture similar distribution in the existing DOE-WQI.