Principal Component Analysis for River Network Data: Use of Spatio-temporal Correlation and Heterogeneous Covariance Structure

Kyusoon Kim

Department of Statistics, Seoul National University

Abstract

Spatio-temporal measurements observed through river networks have two distinct characteristics: a spatio-temporal correlation under the flow-connected structure and the existence of heterogeneous covariances, which require a careful approach to implement principal component analysis (PCA). This presentation focuses on developing a PCA method to reflect the unique characteristics of river networks. We propose a novel method combining flow-directed PCA and geographically weighted PCA for the domain of river networks. The strengths of our approach are that it can (i) reduce dimensionality for streamflow data while effectively removing correlation among them and (ii) identify the group structure of data. It is possible to find essential patterns and sources of variation that may not be disclosed due to the attributes of flow-connected networks. We apply the proposed method to the daily monitoring records of total organic carbon in the Geum River catchment area in South Korea. The results show that the proposed method successfully adjusts for the topological structure of the network and temporal correlation among observations while considering the spatial heterogeneity, enabling a more concrete understanding of monitoring networks.

Keywords: PCA, GWPCA, Flow-connected network, Water quality monitoring, Spatial heterogeneity, CPC