Contribution ID: **39** Type: **not specified**

dPCA: A Python Library for Dynamic Principal Component Analysis

Wednesday, 13 September 2023 11:45 (20 minutes)

Analysis of dynamical systems often entails considering lagged states in a system which can be identified by heuristics or brute-force for small systems, however for larger and complex plantwide systems these approaches become infeasible. We present the Python package, dPCA, for performing dynamic principal component analysis as described by Vanhatalo et al.. Autocorrelation and partial autocorrelation matrices can be constructed for which eigen decomposition can reveal important lags in terms of large eigenvalues and subsequently which variables are highly correlated across time in terms of eigenvector coefficients.
Two use cases are presented –one employing synthetic timeseries data to demonstrate a direct connection to ARMA systems, and one employing two datasets from the largest industrial wastewater treatment plant in Northern Europe. The second use case demonstrates a low-cost tool for analysing large system dynamics which can be used for initial feature engineering for supervised prediction tasks at the plant. The two datasets present different plant layouts utilising different flow schemes, and the approach and Python package is then used to find delays between upstream production plants and downstream operations.

Finally, a perspective is given on how the package can be applied for identifying which lags to use for statistical process monitoring as well as future work.

E. Vanhatalo, M. Kulahci and B. Bergquist, On the structure of dynamic principal component analysis used in statistical process monitoring, Chemometrics and Intelligent Laboratory Systems. 167 (2017) 1-11. https://doi.org/10.1016/j.chemolab.2017

Keywords

Software, Python, Time-series

Classification

Mainly application

Primary authors: TOPALIAN, Sebastian (Technical University of Denmark); KULAHCI, murat (DTU)

Presenter: TOPALIAN, Sebastian (Technical University of Denmark)

Session Classification: CONTRIBUTED Data Science