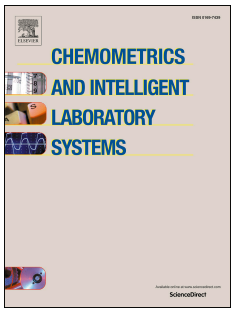
Accepted Manuscript



Comparison of dimensionality assessment methods in Principal Component Analysis based on permutation tests

Raffaele Vitale, Edoardo Saccenti

PII: S0169-7439(17)30804-3

DOI: 10.1016/j.chemolab.2018.08.008

Reference: CHEMOM 3668

To appear in: *Chemometrics and Intelligent Laboratory Systems*

Received Date: 25 December 2017

Revised Date: 4 July 2018

Accepted Date: 12 August 2018

Please cite this article as: R. Vitale, E. Saccenti, Comparison of dimensionality assessment methods in Principal Component Analysis based on permutation tests, *Chemometrics and Intelligent Laboratory Systems* (2018), doi: 10.1016/j.chemolab.2018.08.008.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Comparison of dimensionality assessment methods in Principal Component Analysis based on permutation tests

Raffaele Vitale[1,2,**]

Katholieke Universiteit Leuven & Université de Lille

Edoardo Saccenti[3,*]

Wageningen University & Research

Abstract

We compare the performance of several data permutation methods for assessing dimensionality in Principal Component Analysis. We consider the classical Horn's Parallel Analysis, Dray's approach based on the similarity between the data matrix under study and its lower rank approximation and Vitale *et al.*'s method based on sequential deflation and rank reduction. Their potential is assessed on a large array of simulated data sets accounting for different data correlation structures, data distributions and homo- and heteroscedastic noise, and on 15 experimental data sets from different disciplines, such as metabolomics, proteomics, chemometrics and sensory analysis. In both the simulated and real life case-studies we report differential behaviours of the concerned techniques for which we propose theoretical explanations. The paper also discusses their limits of applicability and some guidelines are offered to the practitioners.

Keywords: Eigenanalysis, Rank approximation, Monte Carlo simulations, Parallel Analysis, Deflation;

1. Introduction

Principal Component Analysis (PCA) [1, 2] is probably the most used tool for the exploration and the compression of high dimensional data sets in many

Submitted to Chemometrics and Intelligent Laboratory Systems *August 13, 2018*

*Principal Corresponding author
**Corresponding author
Email addresses: rvitale86@gmail.com (Raffaele Vitale), esaccenti@gmail.com (Edoardo Saccenti)
[1]Molecular Imaging and Photonics Unit, Department of Chemistry, Katholieke Universiteit Leuven, Celestijnenlaan 200F, B-3001, Leuven, Belgium
[2]Laboratoire de Spectrochimie Infrarouge et Raman - UMR 8516, Université de Lille - Sciences et Technologies, Bâtiment C5, 59655, Villeneuve d'Ascq, France
[3]Laboratory of Systems and Synthetic Biology, Wageningen University & Research, Stippeneng 4, 6708 WE Wageningen, the Netherlands

Download English Version:

https://daneshyari.com/en/article/9953322

Download Persian Version:

https://daneshyari.com/article/9953322

Daneshyari.com