ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/317752485

Visualizing hyperspectral data with linear and non-linear dimensionality reduction methods

Poster · June 2017
DOI: 10.13140/RG.2.2.23506.76485

CITATIONS
0

READS
561

3 authors, including:

Puneet Mishra
University of Strathclyde
40 PUBLICATIONS **140** CITATIONS
SEE PROFILE

Julian Morris
Newcastle University
393 PUBLICATIONS **6,913** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Dynamic Process Performance Monitoring and Spectral & Process Data Fusion View project

Project Multi and hiperspectral images for agricultural products View project

All content following this page was uploaded by Puneet Mishra on 22 June 2017.

The user has requested enhancement of the downloaded file.

Visualizing hyperspectral data with linear and non-linear dimensionality reduction methods

Puneet Mishra[1], Alison Nordon[1], Julian Morris[1]

1- Centre for Process Analytics and Control Technology, University of Strathclyde, Glasgow (puneet.mishra@strath.ac.uk)

University of Strathclyde Glasgow

Introduction

- **Hyperspectral (HS) data** is characterised as high dimensional with a large number of correlated dimensions leading to **a lower intrinsic dimensionality**
- **To visualize**, the HS data is transformed to a **lower dimensionality** explaining the data in terms of clusters
- The **shape of the data manifold** in the higher dimension largely determines the **effectiveness of the visualization method**
- In cases data **manifold is line**ar, classical methods like principal component analysis **(PCA)** and multi-dimensional scaling **(MDS)** can perform well in preserving the structure
- However, when the points in high dimensional space lie very near or in **non-linear manifold**, methods like **PCA and MDS fails** to capture the structure
- This is because the aim of these methods is **to keep distant points as far as possible** and do not utilize information from **neighbouring data points** [1]
- In this case, **non-linear methods utilizing neighbourhood** information can perform better [2].

Objective

- To test the potential of **linear and non-linear data visualisation** technique for visualizing near infrared HS data of tea products.
- Tested techniques: PCA, MDS, Isometric Mapping (ISOMAP) and t-distributed Stochastic Neighbour Embedding (t-SNE)

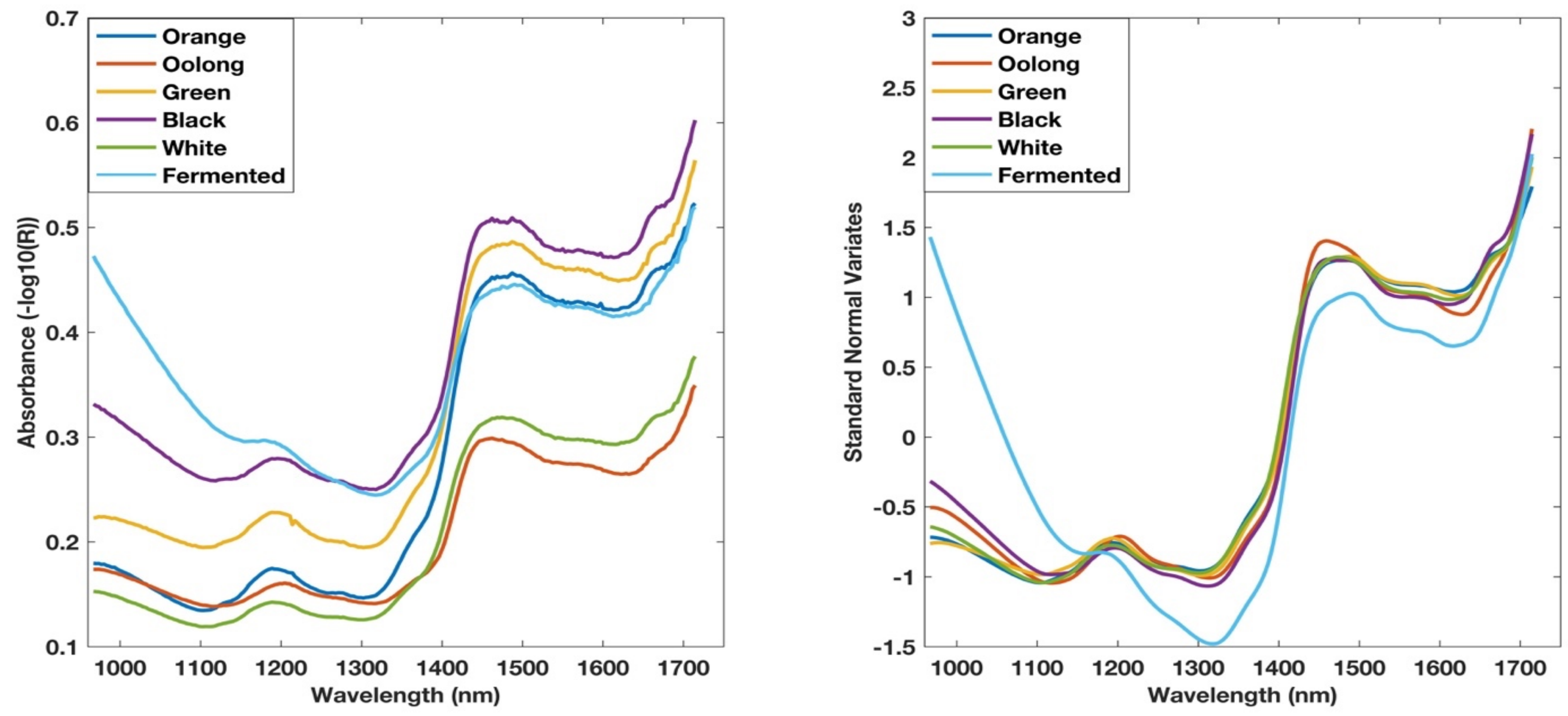
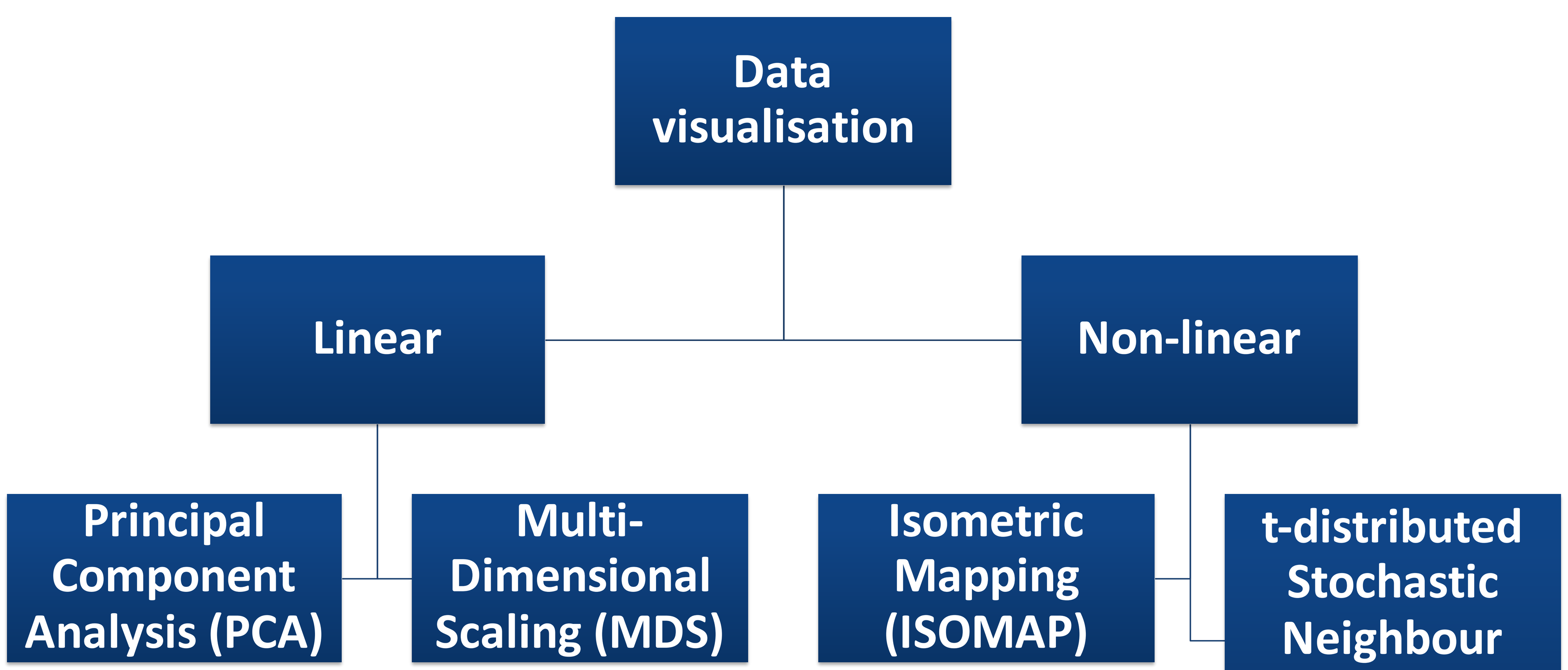


Figure 1: Mean absorbance profile of six different tea samples before and after pre-processing.

Material and methods

- NIRS (950-1750 nm) hyperspectral data of six commercial tea samples (150 spectra each)
    - **Green, Black, Orange, Pu-erh, White and Oolong**
- Data was pre-processed with:
    - Standard Normal Variate
    - Savitzky-Golay smoothing



$min\|TW^T - T_rW_r^T\|^2$

(Score and loadings)

$\min \sum_i \sum_j \|d_{ij} - \hat{d}_{ij}\|^2$

(Euclidean distance)

$\min \sum_i \sum_j \|D_G - D_E\|^2$

(Geodesic distance)

$KL(P\|Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}}$

(Kullback-leibler)

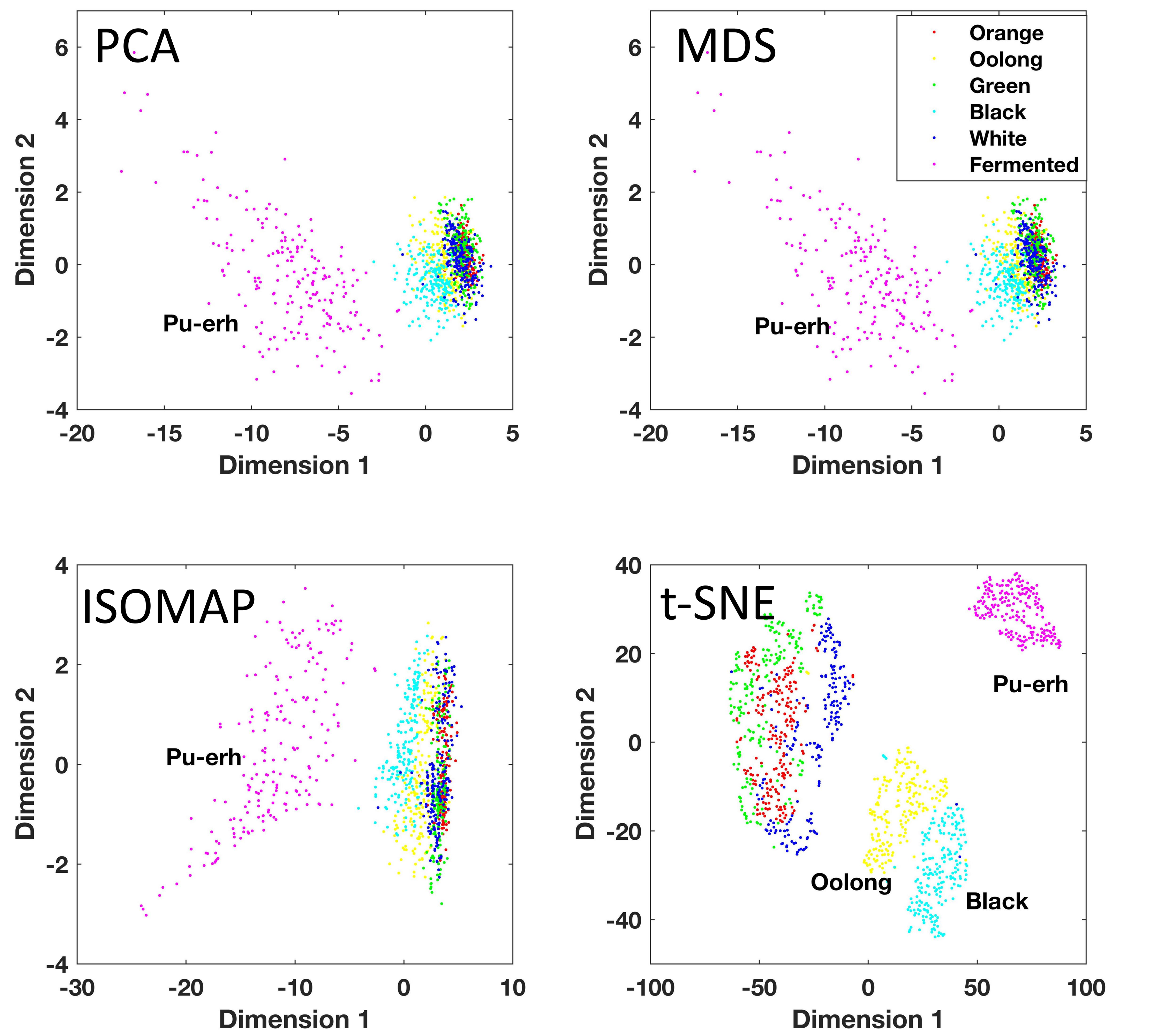


Figure 2: 2D scatter plots for 900 spectra corresponding to six commercial tea samples.

Conclusions

- **Oolong, green, yellow, white, black tea has very similar NIRS profile thus making it difficult for linear methods to separate them**
- PCA, MDS and ISOMAP were able to separate the pu-erh tea from **Oolong, green, yellow, white, black tea**
- t-SNE provided the identification of three different clusters i.e. fermented tea products, oxidised tea products and minimal process tea products.

References

[1] Van Der Maaten, L., Postma, E. and Van den Herik, J., 2009. Dimensionality reduction: a comparative. *Journal of Machine Learning Research*, *10*, pp.66-71.

[2] Maaten, L.V.D. and Hinton, G., 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research*, *9*(Nov), pp.2579-2605.

Mod Life

This work has received funding from the European Unions Horizon 2020 research and innovation programme name "MODLIFE" under the Marie Sklodowska-Curie grant agreement number 675251

MARIE CURIE ACTIONS