Chemometrics for analytical data mining in separation process design for recovery of artemisinin from Artemisia annua

Malwade, Chandrakant R.; Rong, Ben-Guang; Qu, Haiyan; Christensen, Lars Porskjær

Publication date:
2013

Document Version
Publisher's PDF, also known as Version of record

Chemometrics for analytical data mining in separation process design for recovery of artemisinin from *Artemisia annua*

Chandrakant R. Malwade, Ben-Guang Rong, Haiyan Qu, Lars P. Christensen

Department of Chemical Engineering, Biotechnology and Environmental Technology

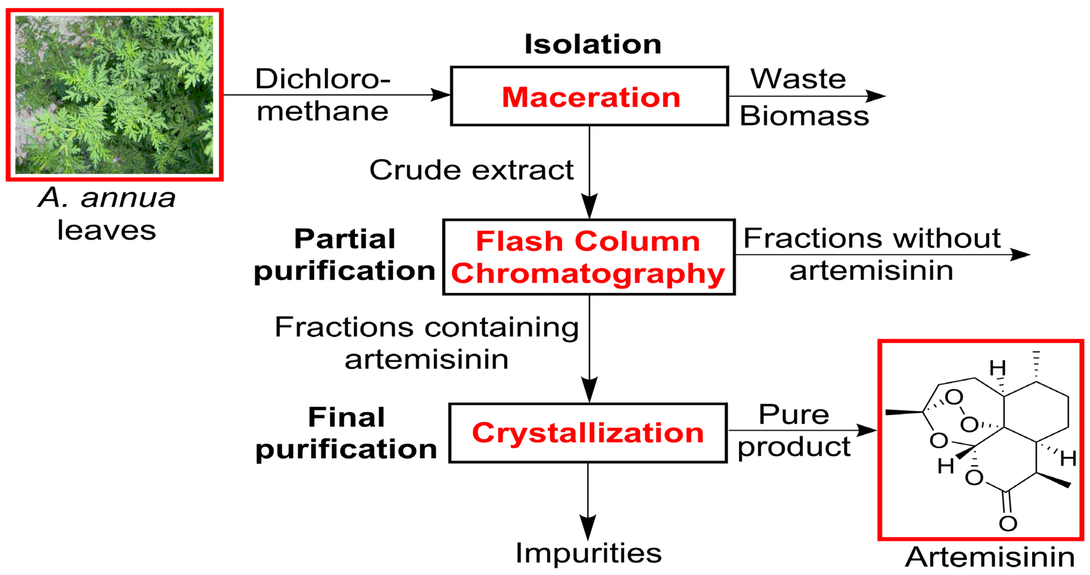
University of Southern Denmark, Niels Bohrs Allé 1, DK-5230 Odense M, Denmark

ESCAPE 23

European Symposium on Computer-Aided Process Engineering

INTRODUCTION

Designing an efficient separation process for recovery of natural products is a challenging task due to the presence of many unknown impurities and lack of fundamental process data. Therefore, rigorous use of process analytical technology (PAT) for qualitative as well as quantitative analysis of raw materials, process streams and final products during conceptual process design stage is inevitable. However, intensive use of PAT often generates enormous amount of data containing many variables which is often difficult to interpret and it is time consuming. In the present work, application of chemometric methods such as PARAFAC to mine the relevant chemical information such as pure elution profiles, UV spectra and concentrations of desired compounds rapidly from the complex HPLC-DAD data is demonstrated.



HPLC-DAD data from 9 fractions (23–31) containing artemisinin obatained from flash column chromatography step (shown above in figure) is used to apply chemometric method PARAFAC.

METHODOLOGY

Chemical analysis
(HPLC-DAD, LC-MS, GC-MS, FTIR, Raman spectroscopy etc)

↓

Data pre-processing
(background correction, noise reduction, baseline correction, shift alignement, normalization etc)

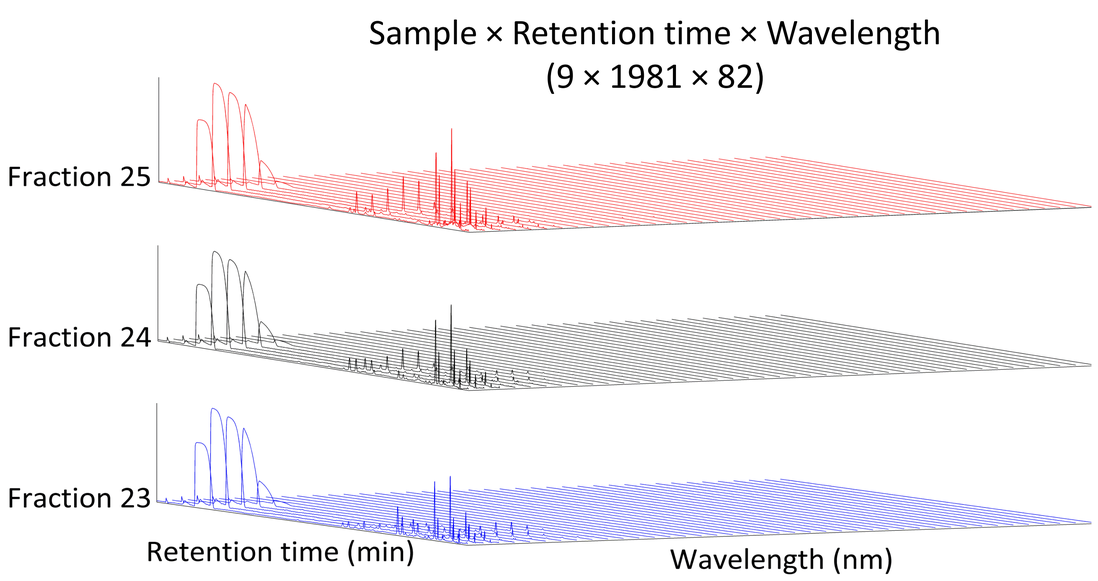
↓

Multivariate statistical data analysis
(Multiple regression analysis, analysis of variance, cluster analysis, factor analysis etc

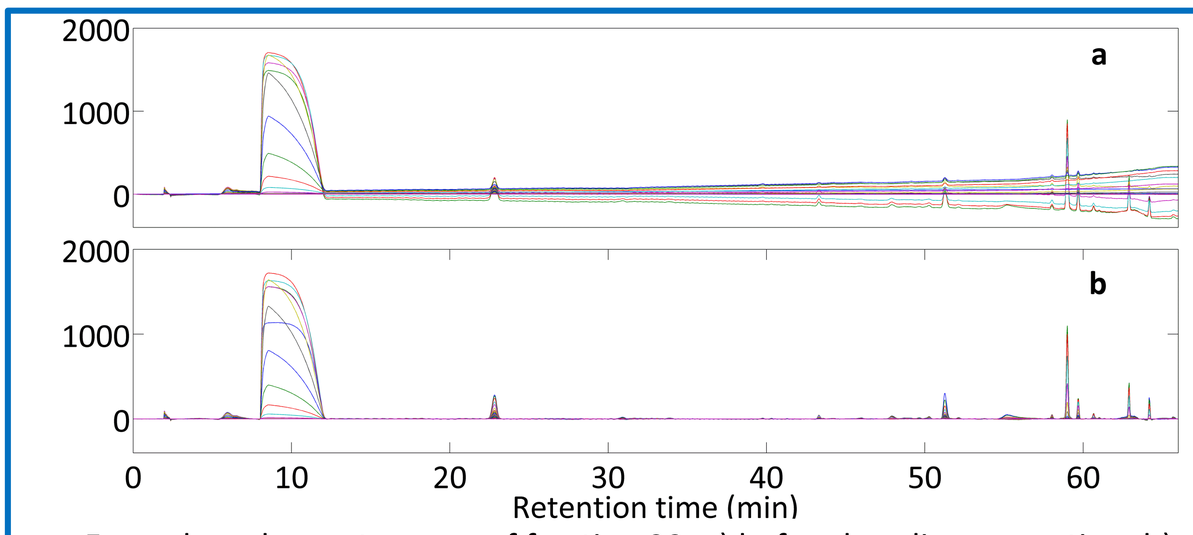
↓

Data evaluation
(Visual inspection of score and loading plots, similarity indices, statistical tests etc)

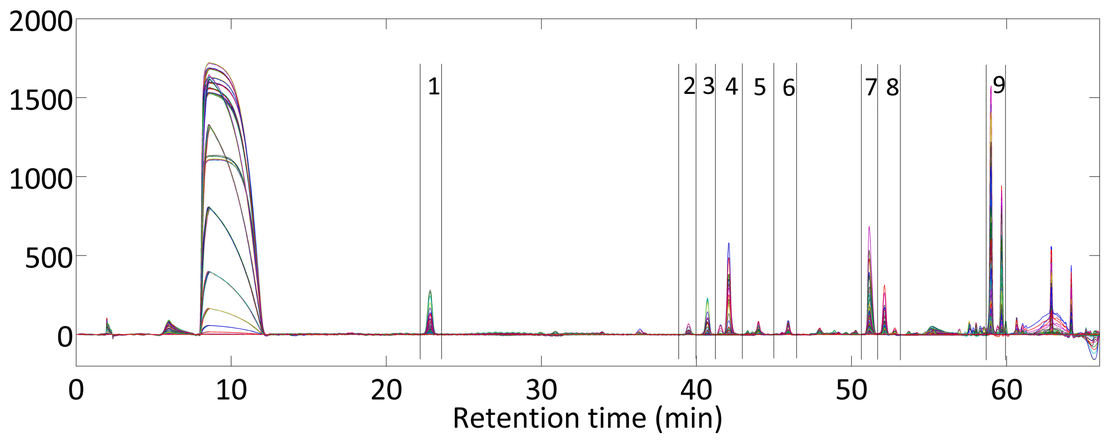
DATASET FROM CHEMICAL ANALYSIS



PRE-PROCESSESING OF DATA

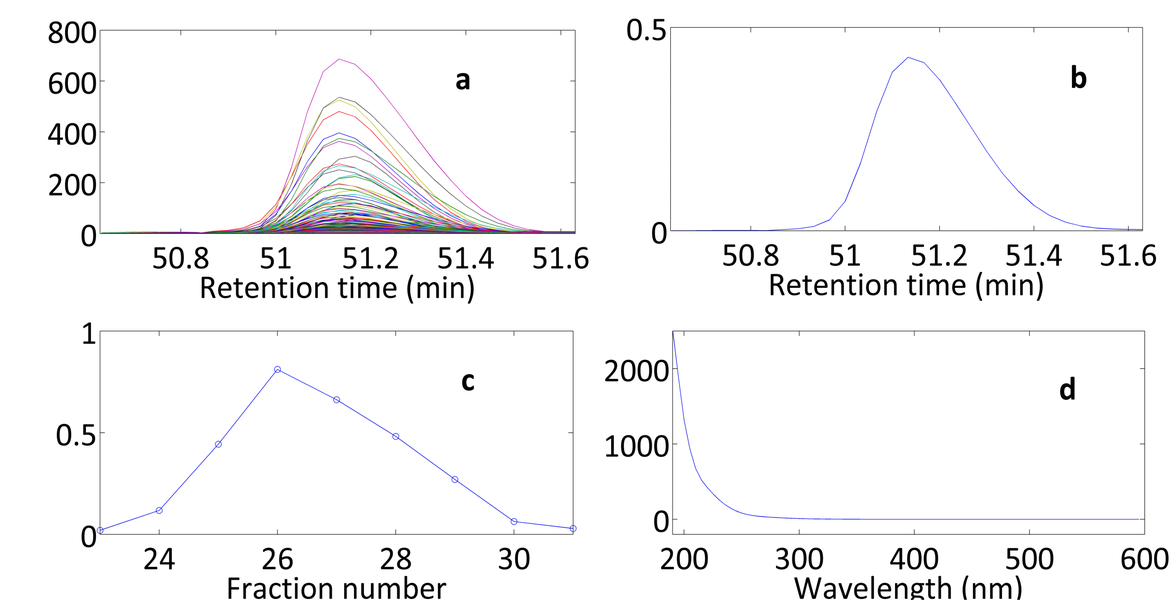


Exemplary chromatograms of fraction 29. a) before baseline correction; b) after baseline correction



Retention time shift alignment of the chemical signals of interest split into intervals. *i*coshift method is used to align the data.

RESULTS – PARAFAC ANALYSIS



One component PARAFAC model fitted to interval containing signal for artemisinin i.e. interval 7. a) Raw data; b) Retention time mode loadings obtained from the fitted model; c) Relative concentration profile of artemisinin; d) UV spectral mode loadings.

Interval Number	Retention time (min)	Number of components	Explained variance (%)	Fractions containing these components
1	22.3 – 23.3	1	98.77	27 – 31
2	38.96 – 40.03	1	99.14	24 – 29
3	40.06 – 41.13	1	98.82	23 – 25
4	41.16 – 42.8	2	99.34	23 – 24
5	42.83 – 44.63	2	95.11	23 – 26
6	44.66 – 46.63	2	95.68	24 – 26 & 29 – 31
7	50.63 – 51.63	1	99.23	23 – 31
8	51.66 – 53.3	1	94.63	23 – 31
9	58.63 – 60.13	2	99.06	23 – 31 & 23 – 27

CONCLUSION

Fundamental process data such as number of components present in the chromatography fractions, their relative concentration profiles, pure elution and UV profiles is successfully obtained from complex analytical chromatography data by applying mutlivariate data analysis technique PARAFAC.