Analytica Chimica Acta 902 (2016) 43–49

Contents lists available at ScienceDirect

Analytica Chimica Acta

journal homepage: www.elsevier.com/locate/aca

Determination of benzo[a]pyrene in cigarette mainstream smoke by using mid-infrared spectroscopy associated with a novel chemometric algorithm

Yan Zhang[a], Hong-Yan Zou[b], Pei Shi[a], Qin Yang[a], Li-Juan Tang[a,**], Jian-Hui Jiang[a], Hai-Long Wu[a], Ru-Qin Yu[a,*]

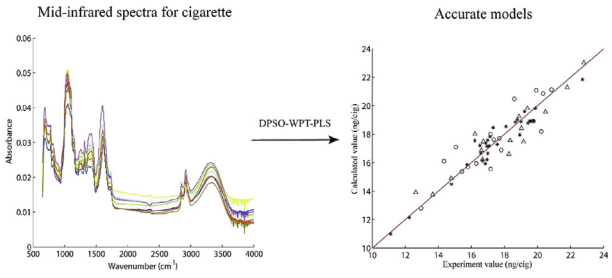
[a] State Key Laboratory of Chemo/Biosensing and Chemometrics, College of Chemistry and Chemical Engineering, Hunan University, Changsha 410082, PR China

[b] Key Laboratory of Luminescent and Real-Time Analytical Chemistry (Southwest University), Ministry of Education, College of Chemistry and Chemical Engineering, Southwest University, Chongqing 400715, PR China

HIGHLIGHTS

- MIR spectroscopy was used to analyze benzo[a]pyrene (BaP) in cigarette smoke.
- A novel adaptive parameter-free algorithm named DPSO-WPT-PLS approach was proposed.
- DPSO-WPT-PLS models the MIR of cigarette samples with promising results.
- A simple, fast, and inexpensive analysis of BaP in cigarette smoke is developed.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 21 October 2015
Accepted 30 October 2015
Available online 4 November 2015

Keywords:
Benzo[a]pyrene
Wavelet packet transform
Particle swarm optimization
Partial least squares
Mid-infrared spectroscopy
Cigarette smoke

ABSTRACT

Determination of benzo[a]pyrene (BaP) in cigarette smoke can be very important for the tobacco quality control and the assessment of its harm to human health. In this study, mid-infrared spectroscopy (MIR) coupled to chemometric algorithm (DPSO-WPT-PLS), which was based on the wavelet packet transform (WPT), discrete particle swarm optimization algorithm (DPSO) and partial least squares regression (PLS), was used to quantify harmful ingredient benzo[a]pyrene in the cigarette mainstream smoke with promising result. Furthermore, the proposed method provided better performance compared to several other chemometric models, i.e., PLS, radial basis function-based PLS (RBF-PLS), PLS with stepwise regression variable selection (Stepwise-PLS) as well as WPT-PLS with informative wavelet coefficients selected by correlation coefficient test (rtest-WPT-PLS). It can be expected that the proposed strategy could become a new effective, rapid quantitative analysis technique in analyzing the harmful ingredient BaP in cigarette mainstream smoke.

1. Introduction

Polycyclic aromatic hydrocarbons (PAHs) mainly derived from incomplete combustion and pyrolysis of organic materials are ubiquitous in such combustion products as included in cigarette smoke.

* Corresponding author.
** Corresponding author.
E-mail addresses: tanglijuan@hnu.edu.cn (L.-J. Tang), rqyu@hnu.edu.cn (R.-Q. Yu).

http://dx.doi.org/10.1016/j.aca.2015.10.029
0003-2670/

Many of these species are carcinogenic and mutagenic, among them benzo[a]pyrene (BaP), the most thoroughly studied of all PAHs, rated as a human carcinogen by the International Agency for Research on Cancer [1]. Therefore, the levels of BaP in cigarette smoke should be monitored for the tobacco quality control and the assessment of the harm of BaP to human health.

Many different methods are developed for detection and quantification of BaP in various systems [2–9]. Despite slightly different versions, a generic method typically utilizes either liquid chromatographic (LC) or gas chromatographic (GC) separation followed by diode array detector (DAD) [2], fluorescence [3], flame ionization detector (FID) [4] or mass spectrometry (MS) detection [5,6]. However, these traditional chromatographic methods are often labor intensive and time-consuming, some requiring sophisticated instrumentation. Although researchers have been developing some rapid BaP methods in aqueous solution such as chemometrics-assisted excitation–emission fluorescence spectroscopy [7] and biosensing [8,9], they are not suitable to cigarette samples.

Near or mid Infrared (NIR/MIR) spectroscopic method is simple, economically attractive and with a direct measurement on the sample, it can easily be adapted and used as an automated method in industry, suitable to be implemented for quality assurance and control of raw material or final tobacco products. The main drawback of NIR/MIR is the wide non selective bands of the spectral profile. In order to enhance the predictive ability of multivariate calibration models, the NIR/MIR raw data are often pre-processed for the elimination of irrelevant information prior to calibration. In this study, the wavelet packet transform (WPT) [10] has been applied to MIR reflectance spectra, followed partial least squares regression (PLS) calibration [11] by using discrete particle swarm optimization (DPSO) algorithm [12] to select informative wavelet coefficients that can be used to characterize MIR spectra and simultaneously search the optimal number of latent variables (LVs) used in the PLS model. The direct benefit is that it makes the proposed algorithm an adaptive method for multivariate calibration, without any parameters to be adjusted. As a comparison, other algorithms, i.e., PLS, radial basis function-based PLS (RBF-PLS) [13], PLS with stepwise regression variable selection [14] (Stepwise-PLS) and WPT-PLS with informative wavelet coefficients selected by correlation coefficient test (rtest-WPT-PLS) [15], were also performed for the same data.

2. Theory and algorithm

2.1. Wavelet packet transform (WPT)

Wavelet packet transform (WPT), which was first introduced by Coifman and Wickerhauser [10], is a powerful signal processing technique. It is an important extension of wavelet transform (WT) [16] capable to offer better resolution for decomposition signals than that of the original WT. Like WT, WPT decomposes a signal in time domain into a series of coefficients (namely, approximation coefficients and detail coefficients) in a new time-frequency space. Consequently, some features inconspicuous in the initial domain might become obvious in the new space. The only difference is that part of higher frequencies is also decomposed continuously in WPT. The procedure of WPT consists of a series of successive decompositions of the signal (with length 2^n) into two components: "approximation coefficients" and "detail coefficients", both with a reduced size of 2^{n-j}, where j is the decomposition level. The approximation coefficients and detail coefficients retain the low-frequency and high-frequency content of the signal, respectively. At each level, the input signal is decomposed by low-pass filters to extract the low-frequency components and by high-pass filters to record the high-frequency components for the next scale. The procedure is repeated with sets of low-pass and high-pass filters until a prescribed level j is reached ($j \leq n$). Fig. 1 schematically describes a 3-level WPT-based signal decomposition process from a standpoint of signal processing.

The same WPT procedure can be conveniently implemented for an NIR/MIR spectrum as long as the time domain is replaced by the wavelength domain. Therefore, it is possible for WPT to magnify some feature of the NIR/MIR spectrum as a "microscope". In fact, WPT has been applied to NIR/MIR analysis, such as transfer of calibration model [17], classification [18,19], signal compression and denoising [20], and multivariate calibration [15,21]. In addition, it is worthwhile to mention that these processed coefficients are used as variables directly rather than reconstruct spectra with approximation and detail coefficients in this study, since WPT transforms a signal linearly from its original domain to a new domain without prejudice [15].

2.2. Adaptively partial least squares based on wavelet packet transform optimized by discrete particle swarm optimization (DPSO-WPT-PLS) algorithm

In the wavelet domain, all information of a NIR/MIR spectrum of a multi-component system is transformed into a series of approximation and detail coefficients. Now what we need is to determine which coefficients are related merely to the concentration of a particular component. Such coefficients can be used as suitable variables to build a multivariate calibration model for predicting the concentration of the component concerned. Generally, PLS is applied for multivariate calibration due to its inherent ability to reduce noise and further enhance the performance of the model. However, when PLS is used, a key parameter in the calibration model is the number of latent variables (LVs) employed, and the optimal value of this parameter would depend on which variables have been selected. Intuitively, we can search the optimal variables (coefficients) under a fixed value of the number of LVs. Repeating this step until every possible value of the number of LVs is covered and choosing the best obtained result, the corresponding value of the number of LVs and the variables selected would form the final solution. However, this brute-force way based on exhaustive search needs a lot of computing power and a lot of time to complete. It is desirable to implement variable selection procedure with the optimization of the number of LVs incorporated in it. For this purpose, a global stochastic optimization technique is needed. Discrete Particle Swarm Optimization algorithm (DPSO) is invoked to select the suitable variables (coefficients) and the proper number of PLS components simultaneously by minimizing the following objective function:

$$\mathrm{Re} = \sqrt{\frac{\mathrm{RSSC} + \mathrm{RSSV}}{I_c + I_v}} \tag{1}$$

where RSSC is the sum of squared of residual of the calibration set, and RSSV is the sum of squared residual of validation set. The former guaranteed the precision of the model; and the latter ensured the generalization ability of the model. I_c and I_v are the number of calibration set and validation set, respectively.

2.3. Discrete particle swarm optimization algorithm (DPSO)

Particle Swarm Optimization algorithm (PSO) [22–24] inspired by social behavior of bird flocking is a global stochastic optimization technique. Like genetic algorithm (GA) [25], PSO is based on population optimization and searches for optima by updating generations. Compared with GA, PSO has no evolution operators and therefore is conceptually simpler and requires lower computation costs [26]. It has been demonstrated that PSO is a robust algorithm