Optik - International Journal for Light and Electron Optics 176 (2019) 600–610

Contents lists available at ScienceDirect

Optik

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

Original research article

# Resolution of overlapping terahertz spectra using non-negative matrix factorization base on pure variables initialization

Jingzhong Gan[a], Binyi Qin[b,*], Yun Li[c], Jie Qiu[a]

[a] School of Computer Science and Engineering, Yulin Normal University, Yulin, Guangxi 537000, China
[b] School of Electronics and Communication Engineering, Yulin Normal University, Yulin, Guangxi 537000, China
[c] College of Chemistry and Food Science, Yulin Normal University, Yulin, Guangxi 537000, China

ARTICLE INFO

*Keywords:*
Terahertz time domain spectroscopy
Resolving spectra
Non-negative matrix factorization
Pure variables initialization

ABSTRACT

In many cases, especially for multi-component mixture, the overlap of spectra will greatly affect the identification accuracy. It is not a reliable method to identify components only by analyzing absorption peaks of mixture directly. Resolution of overlapping THz spectra is a useful way for identification. In this paper, we used the non-negative matrix factorization (NMF) base on pure variables initialization for resolution THz spectra of multi-component mixture. Instead of random non-negative initialization, we introduced the pure variables of which makes the initialization process more intuitive and more interpretable. To verify the proposed method, binary and ternary mixture was preparation. The results show that NMF base on pure variables initialization can resolve the THz spectra of mixture into the spectra of components. Moreover, the initialization method base on pure variables needs fewer iterations than the initialization method base on random non-negative values. It also means that NMF base on pure variables initialization is a potential new approach to identify each pure component in a multi-component mixture.

## 1. Introduction

Terahertz (THz) wave is an electromagnetic wave ranged from 0.1 to 10 THz between microwave and infrared. In contrast to traditional spectroscopic methods, THz response is coupled to rotational and vibrational of molecular, which refers to fingerprints characteristic of the molecule [1]. There are some reports about application of THz spectroscopy for identifying amino acid [2,3], pesticide [4–7] and antibiotic [8–10]. However, these reports are based on a precondition that there is only one component in a mixture to contribute absorption peaks. It is hard to be satisfied in practice. In many cases, especially for multi-component mixture, the overlap of spectra will greatly affect the identification accuracy. It is not a reliable method to identify components only by analyzing absorption peaks of mixture directly. Thus, it is necessary to investigate an approach for extracting more information of components to enhance mixture identification.

Resolution of overlapping THz spectra is a potential way for multi-component mixture identification. But there are rare report on THz spectra resolution. Li reported that the spectra of pure components were extracted from THz spectral of mixtures [11]. However, the method is required a premise that the spectrum of every component should own a distinctive absorption peak. Ma introduced self-modeling technique to resolution of overlapping THz spectra [12]. Nevertheless, the method required user interaction. As we known, THz spectra is a non-negative high dimensional data. Non-negative matrix factorization (NMF) has been proved to be an effective method of approximating high dimensional data where the data are comprised of non-negative components [13–17]. Ma employed

* Corresponding author.
*E-mail address:* qby207@163.com (B. Qin).

https://doi.org/10.1016/j.ijleo.2018.09.094
Received 16 July 2018; Accepted 18 September 2018

NMF for extracting the spectra of pure components from spectral data of mixtures [18]. He selected random non-negative initialization for NMF. But random non-negative initialization is likely to suffer from vast iteration times and lack of physical interpretation of spectra. In order to solve the problems of random non-negative initialization, we introduced the pure variables method to initialize NMF.

In this paper, we used the NMF base on pure variables initialization for resolution THz spectra of multi-component mixture. Pure variables make the initialization process more intuitive and more interpretable, because pure variables are directly proportional to the pure components [19]. In addition, difference value of adjacent eigenvalues was used to estimate the number of components in a mixture before performing NMF. To verify the proposed method, binary and ternary mixture was preparation. And then, correlation coefficient and absorption peaks were used for evaluated the result of NMF. The result indicates that NMF base on pure variables initialization is a potential method to identify each pure component in a multicomponent mixture.

## 2. Experiment

### 2.1. Sample preparation

In this study, a binary and a ternary mixture, which compose two and three components respectively with different weight ratios, were prepared. In the binary mixture, two components, namely imidacloprid and carbendazim, were chosen. In the ternary mixture, three components, i.e. imidacloprid, carbendazim and ascorbic acid, were selected.

Imidacloprid is a systemic insecticide which acts as an insect neurotoxin. It belongs to a class of chemicals called the neonicotinoids which act on the central nervous system of insects, with much lower toxicity to mammals. Carbendazim is a widely used, broad-spectrum benzimidazole fungicide and a metabolite of benomyl. Ascorbic acid is a vitamin found in food and used as a dietary supplement. Carbendazim and ascorbic acid with a purity of 98% was purchased from Adamas Co., Ltd. Polyethylene powder was supplied by Sigma-Aldrich Co., Ltd. Imidacloprid with a purity of 95% was bought from Shanghai Yuanye Bio-Technology Co., Ltd.

For the binary mixture, imidacloprid and carbendazim was mixed with weight ratio ranged from 20% to 90% (w/w). Each sample was prepared with a certain amount of polyethylene (35 mg) as diluent under the pressure of 8 MPa with diameter of 13 mm after drying treatment (423 K) for 1 h in a vacuum drying oven. The total mass of each sample is 185 mg. Table 1 presents the composition of the binary mixture.

For the ternary mixture, imidacloprid, carbendazim and ascorbic acid was mixed with weight ratio ranged from 10% to 70% (w/w). Each sample (total mass 185 mg) was prepared with a certain amount of polyethylene (25 mg) as a diluent under the pressure of 8 MPa. The materials were pressed into tablets with a diameter of 13 mm after drying treatment (423 K) for 1 h in a vacuum drying oven. Table 2 shows the composition of the ternary mixture.

Furthermore, with a certain amount of polyethylene (35 mg) as diluent, pure imidacloprid, carbendazim and ascorbic acid was prepared by the same method mentioned above. The total mass of each sample is 185 mg.

### 2.2. Experimental system

The experimental system of terahertz time-domain spectroscopy (THz-TDS) is composed of two parts: a terahertz time-domain spectrometer Z-3 (Zomega Terahertz Corp., USA) and an ultra-fast fiber laser FemtoFiber pro NIR (TOPTICA Photonics Inc., Germany). The detail of the experimental system has been depicted in previous reports [20,21].

During the experiment, in order to avoid the influence of atmospheric water vapor, the THz beam path was enclosed in a box, and dry air was injected into the box. The relative humidity within the box was kept less than 1%, and the temperature was maintained at room temperature (about 295 K). All the samples were measured in THz transmission mode, and each sample was averaged across every three scans.

**Table 1**
Composition of binary mixture.

| Mixutre no. | Weight ratios (%) | |
|---|---|---|
| | Imidacloprid | Carbendazim |
| sam11 | 20 | 80 |
| sam12 | 30 | 70 |
| sam13 | 40 | 60 |
| sam14 | 50 | 50 |
| sam15 | 60 | 40 |
| sam16 | 70 | 30 |
| sam17 | 80 | 20 |
| sam18 | 90 | 10 |