Chemometrics and Intelligent Laboratory Systems 151 (2016) 136–145

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Hyperspectral image analysis of Raman maps of plant cell walls for blind spectra characterization by nonnegative matrix factorization algorithm

Monika Szymańska-Chargot, Piotr M. Pieczywek *, Monika Chylińska, Artur Zdunek

Institute of Agrophysics, Polish Academy of Sciences, Doświadczalna 4, 20-290 Lublin 27, Poland

ARTICLE INFO

Article history:
Received 8 September 2015
Received in revised form 7 December 2015
Accepted 18 December 2015
Available online 28 December 2015

Keywords:
Chemometry
Raman images
Plant cell wall
K-means
Hierarchical clustering
Nonnegative matrix factorization

ABSTRACT

The aim of this contribution was to develop methods of Raman spectral data analysis with respect to its spatial distribution, produced by a signal deriving complex biological substance. A novel approach based on nonnegative matrix factorization (NMF) combined with the clustering algorithms was introduced for analysis of plant tissue chemical composition. The multivariate approach was tested on the Raman maps of two different tissues of carrot root (*Daucus carota* L. subsp. *Sativus*) — xylem and cambium were captured and analyzed. The initial step of analysis involved pre-processing of individual spectra on two interconnected information levels — spatial and spectral. The proposed approach allowed successful removal of unwanted and corrupted sections of data and replace it with new interpolated values using the nearest neighborhood. The NMF algorithm was tested on refined experimental datasets and showed great performance at reducing the dimensionality of large quantities of spectral information. It also allowed to obtain the pure spectra of individual data components and their concentration profiles which were easily interpretable and had high resemblance to the original data. The output of the NMF analysis was used as a starting point for two clustering algorithms — k-means clustering and hierarchical clustering methods. Both methods converged with similar results providing precise spatial separation of spectral data according to the most predominant component (pectins, cellulose and lignins) in specific area of studied tissues. Obtained clusters distribution showed good match not only with chemical component distribution but also with structural features of tissue samples. Moreover, the proposed method of Raman images analysis allowed to blind spectral separation resulting in rapid and robust analysis of cell wall chemical composition with respect to its spatial distribution.

1. Introduction

The main function of primary cell wall in plants is to maintain the functional integrity during cell morphogenesis and during exposure on biotic/abiotic stress [1]. Moreover, the cell wall has to be strong enough to withstand the turgor pressure and elastic enough to allow the physical enlargement of cells [2,3]. Therefore, the composition of cell wall is vital with the respect to the mechanical properties of plant organs and shows considerable changes during the cell wall growth: newly synthesized polysaccharides are incorporated into the wall and on the other hand action of enzymes could have degraded some of them [4,5]. The cell walls of plant tissue consist of two different phases: the microfibrillar and matrix phase [6]. The matrix phase is made up from a few polysaccharide components, namely pectins, hemicelluloses and also other compounds as proteins and phenolics. The cellulose itself forms a fibrillar phase.

So far, several microscopic methods (transmission electron microscopy TEM, confocal microscopy CLSM, fluorescence microscopy, atomic force microscopy AFM) combined with chemical analysis or immunolocalization have been developed for an examination of the cell wall structure and composition, and thus an observation of dynamic processes occurring therein. However, none of these methods can provide detailed information about the distribution, quantity and structural arrangement of the original building substances of the cell walls with a micrometer spatial resolution. Moreover, this investigation of cell wall composition usually requires the chemical extraction which could lead to changes in structure and properties of isolated polysaccharides or proteins. Therefore there is a demand for the less destructive and non-invasive methods of cell wall composition characterization. In response for these needs confocal Raman micro-spectroscopy has been employed, as a technique that provides information about the chemical composition and distribution of the individual substances in a label-free and non-invasive manner [7–10]. Raman imaging techniques have been successfully applied to observe the differences between various wood tissues such as *Picea mariana* (black spruce) or *Populus trichocarpa* (black cottonwood), or to monitor changes within the distribution of active substances like carotenoids, polyacetylenes or lycopene

Abbreviations: NMF, Nonnegative matrix factorization; SVD, Singular value decomposition.

* Corresponding author. Tel.: +4881 744 50 61; fax: +4881 744 50 67.
E-mail address: p.pieczywek@ipan.lublin.pl (P.M. Pieczywek).

http://dx.doi.org/10.1016/j.chemolab.2015.12.015
0169-7439/

in other plants [8,9,11–15]. The advantage of this technique is that all the polysaccharides are probed at the same time; however, data analysis is considered a very challenging task [16].

Direct imaging of single Raman band intensity [7,17] is one of the simplest and most straightforward methods of acquiring information from data captured by Raman microscopy. Chemical maps are obtained by selection of single band or calculated by integrating over the narrow wavenumber range and are showing spatial distribution of intensities in selected spectral range. Therefore, if selected band refers to vibration of certain chemical bond characteristic for particular chemical substance, the distribution of this component over the sample is obtained. However the single band analysis become somewhat problematic, when sample is composed of multiple compounds with similar chemical composition. Overlapped bands derived from different chemical compounds leads to ambiguous data interpretation where, for instance, similar bands can be assigned to different sources. In such simplified analysis a true chemical composition of sample cannot be fully recognized. On the other hand, the Raman microscopy produces large data sets which can be stored and processed in form of the three dimensional matrices, where rows and columns usually represent planar coordinates of image, while 'depth' of matrix corresponds to spectral information. Processing such large amount of data requires fast computers and large storage capacities and designing new image analysis approaches to handle and extraction of useful data.

In order to improve efficiency of processing and analysis of the multidimensional data sets several algorithms have been already developed and tested: partial least squares (PLS, [18]), principal component analysis (PCA, [19–21]), independent component analysis (ICA, [22]), singular value decomposition (SVD, [23–25]) or vertex component analysis [20]. The main concept of these methods is to estimate the number of more significant components of data set using maximum variance criterion and to remove noise and redundant parts of data. This results in reduction of data set dimensionality without a significant loss of relevant information. The major limitation of these methods is that they do not provide the true signatures and relative concentration profiles of chemical components of image, as they allow resulting matrices to have negative values.

Based on this observation a family of new multivariate methods spawned under common name — self-modeling curve resolution methods. Currently several algorithm are available for different chemometric applications — evolving factor analysis and exhaustive evolving factor analysis (EFA, E-EFA, [26,27]), window factor analysis (WFA, [28]), positive matrix factorization (PMF, [29]), nonnegative matrix factorization (NFM, [30]) and multivariate curve resolution by alternating least squares (MCR-ALS, [19,31–34]). The modeling process in these techniques is carried out under assumptions of a number of constraints, for instance non-negativity, unimodality or equality.

In this paper we explore the capabilities of the nonnegative matrix factorization algorithm (NMF) in unmixing Raman spectral data. The NMF provides blind positive spectra separation by decomposition of the data matrix into the product of two meaningful matrices in that way the remaining data maximizes the variance of processed data set. NMF gives the resolution of complex mixtures where no or little prior information is available. The non-negativity constraint ensures that the profiles of pure-components have physical meaning and can be interpreted as concentration profiles and spectra. The NMF algorithm already have been used in a broad range of applications such as face recognition algorithms [35,36], separation of analytes in nuclear magnetic resonance spectroscopy [37], resolution of graph matching problems [38], and source separation in digital audio signal [39]. In bio-sciences Gobinet et al. [40] applied NMF method to the characterization and fluorescence chemical mapping of wheat grain sections. The NMF provided distribution and concertation of the chemical species which can be used as an indicator of aleuronic contamination of wheat samples. Whereas, Montcuquet et al. [41] tested the NMF algorithm on experimental *in vivo* data and successfully separated signals from fluorescent markers overlapped by autofluorescence signal. Moreover, Li et al. [42] introduced nonnegative matrix factorization with orthogonality constraints and demonstrated its effectiveness in chemical agent detection problem in Raman spectroscopy. Piqueras et al. [34] proposed the use of MCR-ALS scores (concentration profiles) for segmentation purposes of kidney calculi hyperspectral images. K-means clustering was adopted to separate regions of the image based on pixel spectra.

The aim of this work was to test the feasibility of the developed methods of Raman spectral data analysis with respect to its spatial distribution, produced by a signal from complex biological substance. Here we explore the possibilities of application of multivariate methodology and clustering methods in the field of spectral and structural analysis of plant tissues. This study presents a comprehensive approach, including data pretreatment and further steps based on the NMF method to effectively and accurately extract individual components (i.e. different cell wall polysaccharides) and to obtain information on spatial clustering of those substances. The proposed analysis was implemented and demonstrated on cell wall material of two different tissues from carrot root (*Daucus carota* L. subsp. *Sativus*): xylem and cambium. Comparative expert analysis was carried out using single band imaging approach described by Chylinska et al. [7], to show the advantages of proposed methodology in analysis of plant tissue spectral data.

2. Materials and methods

2.1. Material

The carrot root (*D. carota* subsp. *Sativus*) was chosen as a model material for Raman imaging due to well defined structure and complex chemical composition, as well as the availability and ease of preparation. The cell walls of two different tissues of carrot root xylem and cambium were chosen. Xylem is a transport tissue in vascular bundles; usually consisting from tracheids and parenchyma cell types. The tracheids' cell wall are impregnated by lignins. Cambium is layer of actively dividing cells between xylem (wood) and phloem (bast) tissues that is responsible for the secondary growth of stems and roots. The cambium cell walls are composed of cellulose, hemicelluloses and pectins well mixed.

Carrot root was purchased in local grocery and cut into slices using vibratome (Leica VT 1000S). The slices with thickness of 60 μm were placed on a glass microscope slide. In order to avoid as much as possible the unwanted background signal generated for instance by substrate material (glass), slides were covered with aluminum foil on one side as well as tissue slices were soaked in acetone in order to remove plant pigments from tissue [7].

2.2. Raman imaging

The Raman maps were obtained with a DXR Raman microscope (Thermo Scientific Inc., USA), equipped with diode-pumped, solid state (DPSS) green laser $\lambda = 532$ nm with maximum power of 10.0 mW. The spectral resolution of the microscope was equal to 4 cm^{-1} with diffraction grating of 900 lines per millimeters and pinhole aperture of 25 μm. The system uses an air cooled CCD detector. The 20×/0.40 objective was used.

The maps were recorded with spatial resolution of 2 μm in both, x and y, directions. The z direction was fixed during the map recording. The integration time was equal to 8 s and fixed for each scan. A single spectrum in each point was recorded in the range of 3500–150 cm^{-1} of Raman shift for an average of 12 scans. The spectra were not normalized. Two Raman maps were captured with 39×47 sampling points for xylem and 60×31 for cambium tissue.