We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Application of Terahertz Technology in Biomolecular Analysis and Medical Diagnosis

Xin Zhang and Zhuoyong Zhang

Additional information is available at the end of the chapter

Abstract

Terahertz technology is a nondestructive technique, which has progressed significantly in the scientific research and gains highly attention in the analysis of biological molecular, cellular, tissues and organs. In this decade, some studies were reported on the application of terahertz technology in medical testing and diagnosis. Here, we summarize the terahertz characters, terahertz spectroscopy, and terahertz imaging technology combined with chemometrics. This chapter focuses on introducing the research progress on analyzing the tissues of cancers using terahertz spectroscopy and terahertz imaging technology. Furthermore, the problems should be solved, and development directions of terahertz spectroscopy and terahertz imaging technology are discussed.

Keywords: medical diagnosis, medical testing, terahertz imaging, terahertz spectroscopy

1. Introduction

1.1. Development of terahertz spectroscopy in biomolecular detection and medical diagnosis

The terahertz (1 THz is equal to 10^{12} Hz) is an electromagnetic wave located between the infrared and microwave regions of the electromagnetic spectrum. Its frequency is defined from 0.1 up to 10 THz. In this region, the spectra can respond and display spectra absorption to low-frequency vibrational modes of molecules, such as torsional and collective vibrational modes and hydrogen-bond modes, and rotational modes [1, 2]. THz waves have a feature of low energy, non-ionizing which provide the advantage of harmless to analyze living tissues [3, 4]. THz spectroscopic methods have been used in the biological sciences for investigating DNA [5], proteins [6, 7], and tissues [8].

The characterization and quantification of DNA are often regarded as a complex laborious process in bioscience. A number of different techniques are therefore offering a variety of approaches, like spectrophotometry [9], UV-induced fluorescence [10, 11], chip-based nucleic acid analysis system [12], etc., for charactering DNA. The terahertz spectrum has been determined to be a promising candidate for the characterization of DNA. Several alternative methods, including fluorescent chromophore labeling and techniques that use terahertz radiation, have been proposed and are currently in use [13]. Terahertz spectroscopy can characteristic DNA samples pretreatment free, fast, and sensitively. Nagel et al. reported a promising approach for the label-free analysis of DNA molecules using direct probing of the binding state of DNA with terahertz spectroscopy [14]. In comparison with free-space detection scheme formerly used, this method provides an impressively promoted sensitivity enabling analysis down to femtomol levels. Debanjan Polley et al. reported a dielectric relaxation study using terahertz spectroscopy of extended hydration sheathe of dilute aqueous solution of salmon sperm (SS) and calf thymus (CT) DNA samples, which are always used as model organism [15]. They fitted the frequency-dependent complex dielectric response according to a Debye relaxation model, which assumes three relaxation modes in their work on SS DNA and CT DNA. The observed relaxation time constants have high relation with that of bulk water and vary from any particular trend indicating to the extended hydrogen-bonded network of DNA in marginal modification. Though a variety of methods were established for characterization of DNA, they have disadvantages like alteration to the nucleic acid sequence, requirement of a thick DNA testing layer, and conductor structure complexity. From that point, THz spectroscopy has its advantage to be applied on the area of DNA detection.

The THz frequency also corresponds to global correlated protein motions, molecular interaction between protein molecules, which were proposed to be essential to functional conformational changes. Niessen used THz microscopy to inhibitor binding sensitivity and test reproducibility of the narrow-band resonances for lysozyme protein crystals. To achieve the data analysis of THz spectra data, they applied a rapid data acquisition technique. The THz spectra were changed dramatically and can be reproducible with inhibitor binding [16]. Chen et al. proposed an approach for automatic identification of biomolecule terahertz (THz) spectra based on the most used chemometric methods, like principal component analysis (PCA) and fuzzy pattern recognition [17]. Chen investigated THz transmittance spectra of saccharide biomolecular samples, and some typical amino acid and their results demonstrate that THz spectroscopy can be utilized for identification of biomolecules efficiently.

In the application of medical testing and diagnosis, THz spectroscopy and THz imaging have been applied to complex analysis [18]. THz imaging has been used for detecting micrometastatic foci of early-stage cervical cancer in the lymph nodes [19]. The in vivo tissue spectroscopic response in vivo tissues is highly depending on the constituent materials and their physical arrangement for the heterogeneous of tissues. This means that the measurements of in vivo tissues will be different from spectroscopic measurements on homogeneous samples of DNA, saccharide, fat, or proteins. THz interactions with biological components of tissue were reviewed by Smye et al. [20]. Woodward et al. demonstrated the application of terahertz pulse imaging (TPI) on skin or related cancer tissues. Using this technique, they detected imaging in reflection approach for the study of skin tissue and corresponding cancer tissue both in vitro and in vivo. The sensitivity of terahertz radiation to polar molecules makes THz spectroscopy and imaging be used for analyzing the hydration levels in the skin. And, it also has potential to

be applied on the preoperative determination of the lateral spread of skin cancer. The terahertz pulse shape in the time domain was studied, and the results show that they were able to differentiate diseased and normal tissues for the study of basal cell carcinoma [21].

The study of Nazarov et al. showed that in the terahertz frequency range, small organic molecules had characteristic absorption lines. Large molecules and tissues had crucial absorption linearly increases with frequency. THz refection spectroscopy provides possibility to study strongly absorbing substances [22]. Using differences in THz absorbance, pathologic diagnosis between normal and cancerous tissues has been reported in the medical literatures [23, 24]. Knobloch's results show that different kinds of tissues can be clearly distinguished from both the larynx of a pig and cancerous human liver using THz spectroscopy. Cherkasova et al. studied human and rat skin reflection spectra in vivo and the effect of glucose and glycerol on these spectra by THz spectroscopy [25]. Variations in skin optical properties were found in the frequency of 0.1 THz.

Hyperspectral imaging can provide information in both space and spectral dimensions. THz imaging has potential for tissue analysis and medical diagnosis as a very promising harmless approach for future imaging applications. It is hopefully to develop the THz imaging system in the future, which requires high-frequency resolution and a cost-effective and much more compact setup that does not necessarily require a laboratory environment based on the development of THz techniques.

1.2. Application of chemometrics on terahertz spectra analysis

THz spectroscopy should take advantage of chemometrics, which has applied in other fields of spectroscopy like infrared, near-infrared, Raman, or fluorescence. Chemometrics [26] provides multivariate tools for exploring the relationships among the objects and tested variables in collected dataset as well as classifiers. Chemometrics has been applied for qualitative and quantitative analysis of the THz spectra of not very complex mixture systems. Absorption intensity of THz spectra is proportional to the concentration of analytes in commonly used dynamic ranges; thus, normal linear modeling methods in chemometrics can be used in their calibration and prediction [27].

A review of terahertz pulsed spectroscopy summarized the most common chemometric methods applied for processing the THz spectra including the way of quantitative univariate and multivariate methods, and it can be found from this chapter [28]. Two quantitative analysis approaches are mostly used: the first is the method applied to a single spectrum without any calibration but using the intensity of the spectra for modeling, and the other is calibration based on a series of THz spectra of reference samples in order to predict quantitative information from unknown samples.

THz spectroscopy combine with chemometrics was reported for quantitative or qualitative analysis of mixture systems in environment, food, agriculture, material, biology, and medicine. Otsuka et al. did quantitative analysis of mefenamic acid polymorphs by terahertz spectroscopy with chemometric methods [29]. They studied the effect of spectra data preprocessing on the chemometric parameters of the calibration models. Hua et al. mostly used regression models like partial least squares (PLS) and principle component regression (PCR) methods for quantitative evaluation of cyfluthrin in n-hexane by THz-TDS [30]. Partial least squares (PLS) is one of the most effective and reliable methods normally being applied for quantitative

analysis of various spectra. El Haddad et al. applied principal component analysis (PCA), PLS, and artificial neural networks (ANN) to quantitative analysis of ternary mixtures by THz-TDS [31], and they obtained good results. Ellrich et al. presented a postscanner by THz spectroscopy using chemometric methods for the evaluation of detected THz fingerprints [32].

Assessment of THz spectroscopy with chemometrics is still under studying including preprocessing, data selection and calibration methods. We strongly believe that THz spectroscopy should take advantage of multivariate analysis for advanced data processing, classification and calibration methods of chemometrics.

2. Instrumentation

For all of the work introduced in this chapter, a transmission THz-TDS cell configuration was used, as depicted in **Figure 1**. **Figure 2** gives schematic of a terahertz time-domain transmission spectrometer system used in this chapter. The THz spectroscopy system was equipped with a commercially available femtosecond laser (SPECIM, MaiTai) for generating the THz pulse. The femtosecond laser light is separated into two beams using a prism. One beam is the probe, and it travels across a free space to focus on the detecting antenna, so that the probe beam provides a relative time delay periodically. The other beam goes through a GaAs-based semiconductor antenna to generate the THz pulse. Then, a parabolic mirror with a hemispherical silicon lens is applied for improving the coupling efficiency of the THz radiation. The beam that passed through the sample placed at the focus of the parabolic mirror, and it is collected by another parabolic mirror. Finally, a photoconductive detector is used for signature collection [33]. In the experiments, the volume of the THz spectra system was filled with dry nitrogen (N_2) to reduce absorption caused by vapor in air.
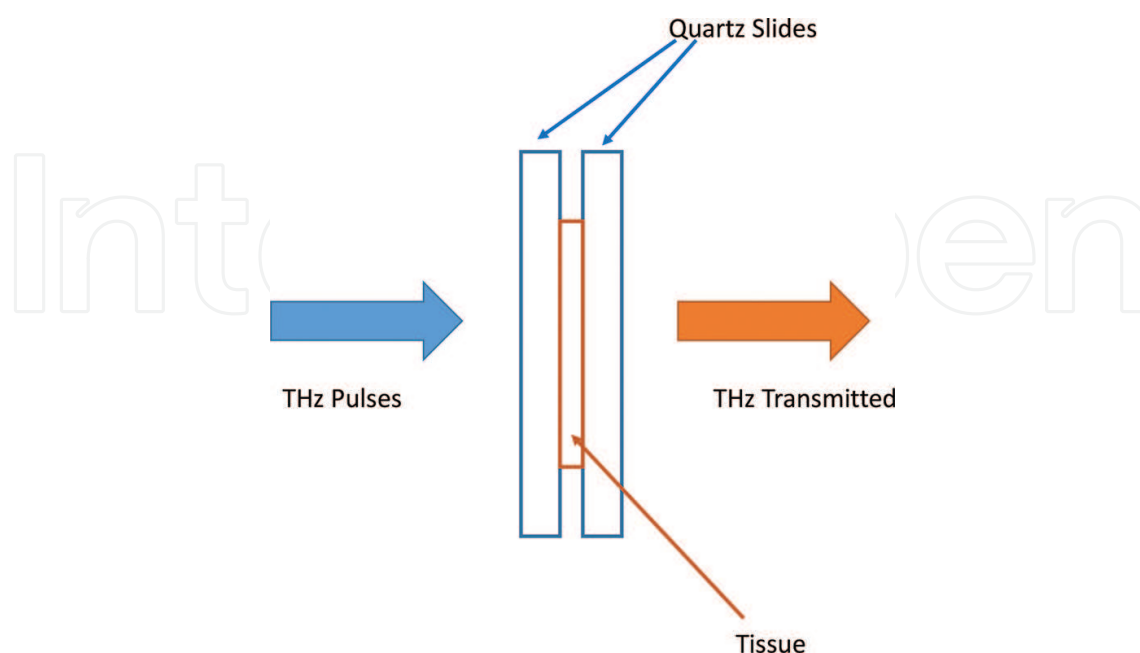
Figure 1. Schematic representation of tissue samples for TDS measurement.
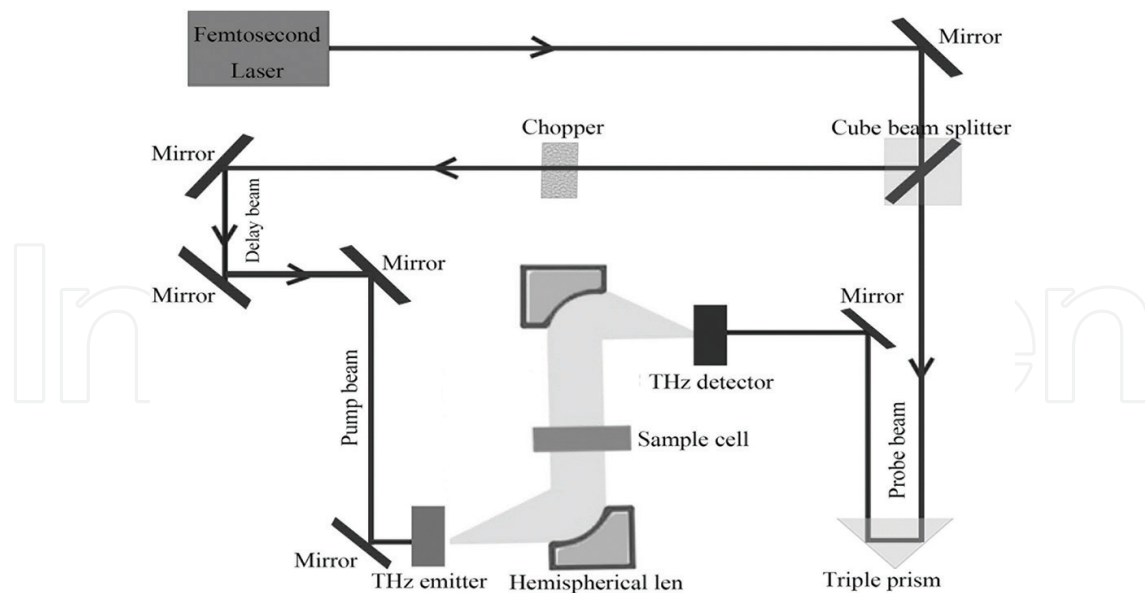
Figure 2. Schematic of a terahertz time-domain transmission spectrometer system used in this chapter.

3. Theory and practical applications

3.1. Parameter extraction from THz spectra

The measurement of the reference pulse and the sample pulse are necessary for calculating the THz absorption coefficient of a sample. When the tissues are analyzed by THz spectra for medical testing or diagnosis, the sample pulse collected is transmitted through the tissue slides, and the reference signal is the THz signal transmitted without the tissue slides. The THz electric field pulses can be calculated as a function of time and the frequency for both signal crossed the tissue sample and reference passed nothing. The frequency domain spectra are obtained by the fast Fourier transform (FFT) in this work. The refractive index $n(\omega)$ describing the dispersion and absorption coefficient $\alpha(\omega)$, describing the absorption characteristics, can be calculated through the following equations [33]:

$$n(\omega) = \frac{\varphi(\omega)c}{\omega d} + 1 \tag{1}$$

$$\alpha(\omega) = \frac{2\kappa(\omega)\omega}{c} = \frac{2}{d}\ln\frac{4n(\omega)}{A(\omega)(n(\omega)+1)^2} \tag{2}$$

for which ω is the frequency, and $\rho(\omega), k(\omega)$ and $\phi(\omega)$ are functions for the amplitude ratio, extinction coefficient, and phase difference of the sample and reference signals, respectively. d is the thickness of sample, and c is the velocity of light in vacuum.

3.2. Chemometric methods applied in this chapter

3.2.1. Data preprocessing methods

The Savitzky-Golay (SG) method is a polynomial filter that performs numerical differentiation and smoothing [34]. This filter removes the noise in the dataset analyzed and simplifies

the computation during the model building. SG method has the ability to process the signals with little delay and with no shifts of the peaks, and it can be performed in a computationally efficient procedure by applying least squares on subsets of the data. In the dataset, a window moves forward within 2m + 1 points fitting by a polynomial of degree p (in which $p \leq 2$ m). The dth ($0 \leq d \leq p$) differentiation of the original data at the midpoint is obtained by performing the fitness polynomial. Finally, the convolution of the entire input data with a digital filter of length 2m + 1 is performed by running least squares polynomial fitting [35, 36].

The multiplicative scatter correction (MSC) is a transformation method used to cope with scaling and offset effects, which is mostly applied in spectral data analysis [37]. It is used to counterbalance for additive and/or multiplicative effects in spectral data. MSC assumed that each spectrum collected from the samples is determined on one hand by the actual sample characteristics and on the other by the particle size. It can also decrease or remove physical effects like particle size and surface blaze, and it corrects differences in the baseline and in the trend. The sample preparation for THz spectra collection always has problems about the difference of particle size during grind and compactness during tableting. So, it has an advantage that the transformed spectra are similar to the original spectra without the effect of baseline effect and the trend of a standard spectrum and that an optical interpretation is therefore more easily accessible [38].

Orthogonal signal correction (OSC) can be applied to remove systematic noise such as baseline variation and multiplicative scatter effects, which is a data processing technique introduced by Fearn [39]. The basic idea of the OSC method is to remove the systematic variations in the collected data that are orthogonal or not related to the properties of the dependent variables. The removed information can be structured noise, such as baseline, instrument variation and measurement conditions. Some reports show that the use of OSC may not result in calibration models with lower prediction errors than models based on raw data. The advantage of using OSC lies in the analysis and interpretation of the corrected data but not in decreased prediction errors. By removing orthogonal information, the important calibration information will be concentrated in fewer principal components instead of being distributed among many linearly dependent variables.

The PC-OSC method used the constraints based on OSC and also applied the theory of principal component analysis (PCA). The detailed procedure of PC-OSC can be referred in Ref. [40]. The emphatic orthogonal signal correction (EOSC) method can be used for the baseline correction of Raman spectra or near- infrared spectra, and it is a method that can be extended to apply on THz spectra. The theory and procedures of EOSC can be found in Refs. [4, 41].

Asymmetric least squares (AsLS) method calculates complex baseline shapes by adjusting the asymmetry parameter and the smoothness parameter. The asymmetry parameter is related to the position of the baseline, and the smoothness parameter related to the flexibility in the shape of the baseline. The AsLS method has been proposed by Zhang et al. [42]. By minimizing the penalized least squares function based on the Whittaker smoother, AsLS method estimates a background contribution and removes or decreases the baseline [43]. The application of this method can be found in Ref. [44]. The initial range of lambda is 10^2–10^5, and p is 0.099, which are experienced parameters based on literatures and our laboratory work experience. We adapted the results to PLS models, respectively, and obtained the best result when lambda is 10^3 [27].

Wavelet transform not only can compress the data to extract feature information but also can remove the noise in the spectral data. In brief, wavelet analysis is based on the wavelet transform of analysis signal. The signal of different spatial scales (frequency) is divided into high-frequency and low-frequency part, and the position of each component on the time axis remains the same. A more detailed description of wavelet analysis algorithm can be found elsewhere [45].

Figure 3a displays raw THz absorption spectra of the prepared samples of binary amino acid mixtures of L-glutamic acid and L-glutamine. **Figure 3b**, **3c** and **3d** displays the THz spectra of the binary amino acid mixtures with preprocessing of SG smoothing, MSC, and AsLS, respectively. From **Figure 3**, we can observe that the absorption spectra with processing of SG smoothing have eliminated the effect of noise and display spectral characteristics more clearly. By using MSC, the scaling and offset effects were removed. Normally, using OSC can get very similar results to that applied in MSC. AsLS can eliminate or decrease the baselines in the THz spectra of the samples. The absorption bands are more easily attributed to certain wavelength in this work. The binary mixtures under different concentration ratios display certain quantitative relation on THz absorption coefficients, which would be the basis for quantitative analysis by THz-TDS transmission spectroscopy. We did not apply wavelet transform on the sample analysis in this chapter, for the other preprocessing could obtain good result already. Anyway, wavelet transform still should be considered in the future work, since it is a powerful method for de-noising, compressing the dataset, removing the background information, etc.
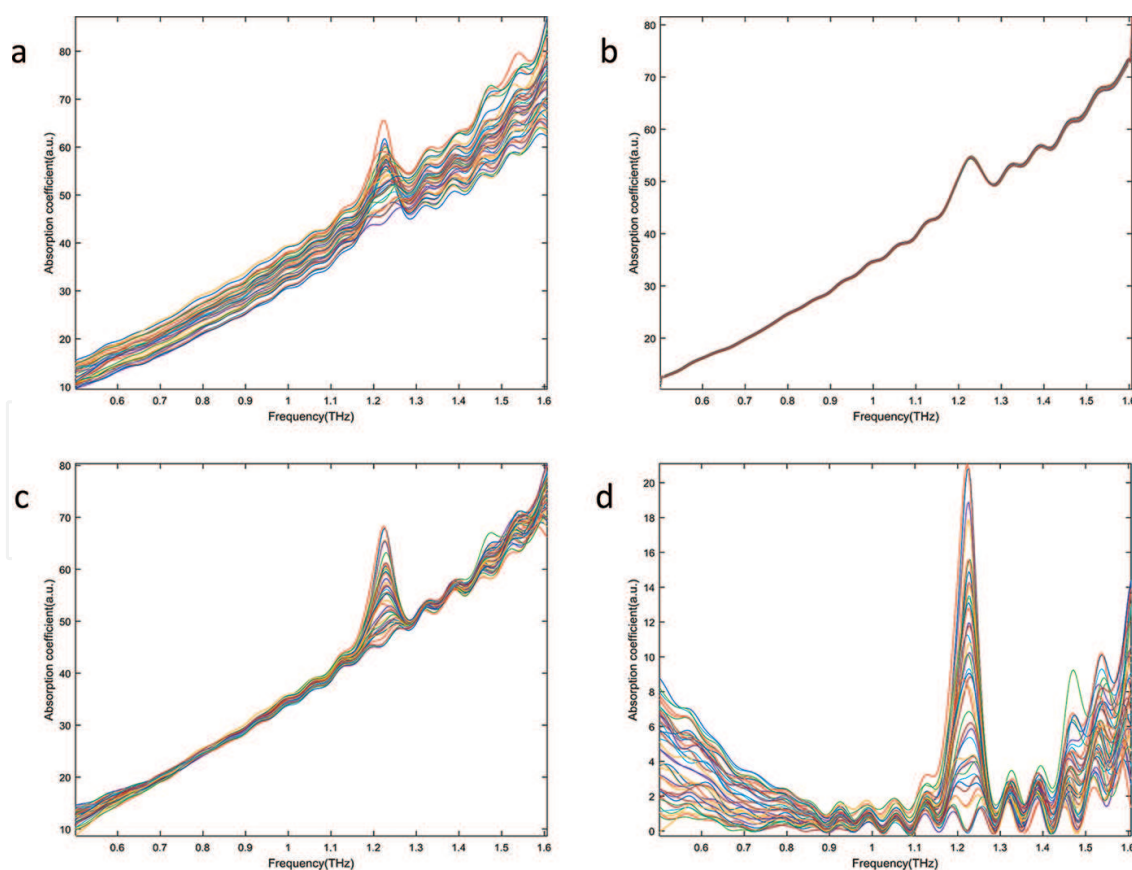
Figure 3. The spectra of the L-glutamic acid and L-glutamine mixture samples after different preprocessing are (a) raw THz spectra, (b) the spectra after smoothing, (c) the spectra after MSC, and (d) the spectra after AsLS.

3.2.2. Examples of chemometric methods for regression model building of THz spectra

3.2.2.1. Principal component analysis

Principal component analysis (PCA) is commonly used to reduce the number of predictive variables. PCA condenses all the spectral information into a few linear combinations of the latent variables instead of the original variables. The linear combinations of the variables can be used to summarize the data without losing too much information but remove noise in the process [46]. It can be used to identify the underlying structure of large datasets and can be used to identify groups within the data from complex mixtures. It can also be used for removing any contribution from noise.

3.2.2.2. Partial least squares

Partial least squares (PLS) is one of the most used chemometric techniques to build quantitative model based on principal component analysis and principal component regression. PLS extracts the orthogonal features from the spectrum and then constructs the correlation between the spectra matrix (independent variable, X) and concentration matrix (dependent variable, Y). The detailed procedure of PLS can be referred to in Ref. [47].

We have investigated PLS for quantitative analysis of L-glutamic acid and L-glutamine using THz spectroscopy. Also, we compared the difference between iPLS and PLS. iPLS divided the whole spectrum into several intervals and builds PLS model for each subset to evaluate the most suitable sub-dataset for a stable model. The subset or several subsets with the lowest root mean square error of cross validation (RMSECV) are chosen for the PLS model building [27]. The iPLS yielded better results with low RMSEP (0.39 ± 0.02%, 0.39 ± 0.02%), and higher R^2 values (0.9904, 0.9906) for glutamine and glutamic acid comparing to the conventional PLS models.

We also analyzed binary isomer of saccharide mixtures, D-(−)fructose and D-(+)galactose anhydrous quantitatively using THz-TDS combined with PLS. The result showed that correlation coefficient (R^2) between true and predicted values is higher than 0.9773. The mean value of root mean square error of prediction (RMSEP) in cross validation set was less than 1.26%.

Therefore, THz-TDS combined with chemometrics is feasible for quantitative analysis of the biomolecular mixtures and may also be extended to analysis of more component mixtures.

3.2.3. Examples of chemometric for classification model building of THz spectra

3.2.3.1. Partial least squares-discriminant analysis

Partial least squares-discriminant analysis (PLS-DA) is a supervised classification method based upon partial least squares regression [47]. The PLS-DA algorithm models the relationship between the measured variable of the dataset and the target variables corresponding to the class label [48]. PLS-DA extracts the latent variables by reducing the dimension

of the dataset like principal component analysis and finds the maximum separation among the classes. The latent variables explain both the variance of the THz spectral data and the high correlation with the response matrix that encodes the class membership [33]. Component number, which has high relative with the accuracy rate and the percentage of the explained variable, is a very important parameter for prediction accuracy and explanation of the model that needs to be estimated for PLS-DA.

3.2.3.2. Support vector machine

Support vector machine (SVM) is a powerful machine learning method with associated learning algorithms that analyze data used for classification and regression analysis [49]. For classification tasks, based on the structural risk minimization principle, this method attempts to find the separating hyperplane which has the largest distance from the nearest training data points. LIBSVM was one of the mostly used toolboxes, and the SVM calculations in the example in this chapter used the linear kernel function [50].

The two essential factors that affected SVM classification performance are (1) error penalty parameter C, which is the compromise between the proportion of error classification samples and algorithm complexity, and (2) form of kernel function and its parameters. Different kernel functions have influence on the classification performance, while different parameters of same kernel function may also affect the results.

3.2.3.3. Fuzzy rule-building expert system (FuRES)

FuRES is a classification tree model using fuzzy entropy of classification which each rule is a temperature-constrained sigmoid logistic function [51]. The rules of FuRES are similar to the processing units in most feed-forward artificial neural networks, while FuRES processing units differ in that the weight vector is constrained to an optical length. The inductive dichotomizer 3 is used for classification by minimizing $H(C \mid A)$, the classification entropy. The weight vector w should be normalized before modeling. A temperature parameter t is used to control the fuzziness of each rule. When optimizing the computational temperature, the maximized extend of the entropy for classification can be found [4]. The equations for this method are given below:

$$\chi_A(x_k) = (1 + e^{-(x_k w - a)/t})^{-1} \tag{3}$$

for which a is the bias value and $x_A(x_k)$ is the degree of fuzzy membership of object x_k. [4]. The conditional probability $p(c_i \mid a_j)$ is obtained by summing the membership functions with the attribute a_i and the class of c_i. The equation is given as follows:

$$p\left(c_i \mid a_j\right) = \sum_{k=1}^{n_i} \chi_A(x_k) \Big/ \sum_{k=1}^{n} \chi_A(x_k) \tag{4}$$

where n_i is the number of objects in class c_i. The classification entropy $H(C \mid a_i)$ of the attribute a_i is given by

$$H\left(C \mid a_j\right) = -\sum_{i=1}^{n} p\left(c_i \mid a_j\right) \ln\left[p\left(c_i \mid a_j\right)\right] \tag{5}$$

The classification entropy $H(C \mid A)$ of the system is the weighted sum of the entropy for each attribute:

$$H(C \mid A) = \sum_{j=1}^{2} p(a_j) H\left(C \mid a_j\right) \tag{6}$$

The FuRES model provides inductive logic in the tree structure of the classifier. In this way, it can accommodate overlapping the data and avoid overfitting the data by the temperature constraint. The advantage of fuzzy classification trees comparing to network classifiers is that they furnish a simple inductive structure that is amenable to interpretation.

3.2.3.4. Fuzzy optimal associative memory (FOAM)

An optimal associative memory (OAM) is using a one-way data to replace binary image of encoded multivariate data as a two-way binary image for fuzzy method [52]. Bipolar matrix with similar size grid unit is built first. A vector of v variables is converted to $v \times h$ bipolar matrix. After removing u unused grid, the number of grid is k $((v \times h) - u)$ [4]. The FOAM stores pattern in a weight matrix **W,** which expressed by

$$\mathbf{W} = \sum_{i=1}^{n} \mathbf{y}_i \mathbf{y}_i^{\mathbf{T}} \tag{7}$$

for which $\mathbf{y}_i$ is the ith bipolarly encoded approach [4]. The stored grid-encoded spectra are orthogonalized to form a basis using singular value decomposition. The encoded predicted background scan z_f can be obtained by

$$\mathbf{z}_f = \mathbf{V}\left(\mathbf{V}^{\mathbf{T}} \mathbf{z}_i\right) \tag{8}$$

for which **V** is a matrix following the orthogonalized pattern of the collected data variables. Then, z_f is decoded to a spectrum vector by changing the gridding procedure. The data object can be assigned to the corresponding classes to make the building fit with the minimum error after reconstructing the raw data. In this chapter, FOAM used its standard configuration of 100 intensity grids and a 19-point triangular fuzzy membership function [4].

3.2.3.5. Validation

Bootstrapped Latin partitions (BPLs) are generalized for evaluation of calibration methods based on cross validation and random sampling verification [53]. Unbiased evaluation of classification or calibration methods is important, especially as these methods are applied to increasingly complex datasets that are under-determined like THz spectra dataset. Precision bounds, such as confidence intervals, are required for interpreting any experimental result. By using BPLs unbiased and reliable evaluation can be gotten by systematically model with samples drew from an arbitrary discrete distribution.

3.2.3.6. Example of diagnosis of cervical cancer

In this work, THz-TDS system was applied to detect the normal and malignant tissue sections as an example for chemometric applications on THz spectral analysis. The classification models combined with different pretreatment methods were established to build a new diagnosis

technique for cervical cancer diagnosis based on terahertz spectroscopy. The effects of different preprocessing methods on THz spectra data to de-noise, remove baseline and optimize model were investigated.

The normal and cancerous cervical tissues were collected and provided by Beijing Haidian Maternal & Child Health Hospital. To keep the tissues, all the cervical tissues were put into 4% formaldehyde solution. The tissues were washed with ethanol solutions for dehydration when we analyzed the samples. The tissues were put into xylene for hyalinization and then embedded by paraffin wax before sliced into 8 μm thick sections. The water-flatted sections were spread upon quartz plates and then put in a regulated heating oven and dried. Two replicate slides were prepared for each of the tissue sections [33].

To establish a model for diagnosis of cervical cancer, PLS-DA, FuRES, FOAM, and SVM were used to build classification models. The parameter for FuRES and FOAM was determined using a self-optimizing PLS-DA from the training datasets. A method to verify the accuracy of classification and calibration models, bootstrapped Latin partition, was used for cross validation of the calibration dataset. When we built PLS-DA and SVM models for this study, the matrix of category variables was used, one for the normal samples and two for the cancer tissue samples. When the predicted value of an external test tissue sample in PLS-DA model was smaller than 1.5, the sample was assigned to normal class or assigned to cancer class otherwise [33].

The classification results of these methods were compared when the data was processed with different preprocessing. MSC, SG smoothing, SG first derivative, EOSC, and PC-OSC were used for pretreatment of the THz spectra, respectively, and the data were normalized before modeling. The performance of preprocessing methods applied in this work was compared. The results of the modeling approaches after pretreatment are evaluated by the pooled prediction rates. The raw data were divided as training datasets and prediction sets based on KS method. The pretreatments were constructed from the training datasets and applied to the prediction sets. For cross validation, five Latin partitions bootstrapped 50 times were applied to evaluate the prediction accuracy of the classification models with different parameters and different pretreatments. For each bootstrap, the data were separate as training set and test set, and each spectrum was used only once in the test set. Each time, four Latin partitions were used for calibration during model building, and the fifth was used for prediction. The predicted results of the five test sets from each partition were pooled back after each validation was finished. This approach was used for evaluations of all the classification models in this work. The average prediction results were calculated with 50 bootstraps to give 95% confidence intervals [33]. The number of components of the OSC model was selected by finding the maximum average classification rate across internal 100×5 bootstrap Latin partitions. All model optimization and construction were performed in MATLAB (MATLAB 7.14.0.334. The MathWorks Inc.).

The obtained results showed that the FuRES and FOAM model using Savitzky-Golay smooth by the first derivative and PC-OSC as pretreatment methods had provided a good predictive results and their classification rates are 92.9 ± 0.4 and 92.5 ± 0.4 %, respectively [4]. The results of the proposed methods show that terahertz spectroscopy combined with fuzzy classifiers

could supply a technology which has potential for diagnosis of cancerous tissue. Combining SG first derivative with PC-OSC as signal pretreatment procedure, the prediction accuracies of the optimal SVM and PLS-DA were 94.0 ± 0.4 and 94.0 ± 0.5%, respectively. Therefore, SVM and PLS-DA with the combination of SG first derivative and PC-OSC based on terahertz spectroscopy of tissue can also provide a good application for diagnosis of cervical carcinoma.

Comparing the classification accuracies pretreated by different preprocessing methods, it indicated that the classification models applied based on terahertz spectroscopy of tissue could provide a better application for early diagnosis of cervical carcinoma, with high classification accuracies. Coupled with terahertz technology, the proposed procedure could provide a convenient, solvent-free, and environmentally friendly application that had a potential development as cancer diagnosis method.

3.2.4. Examples of chemometric for resolution of THz spectra data

3.2.4.1. Multivariate curve resolution

Multivariate curve resolution (MCR) is designed to solve the analysis problem of mixture systems following bilinear model. The MCR methods decompose the raw mixed measurement datasets into matrix corresponding to pure concentration profiles and pure spectra. Constrains following physical and chemical property can be flexibly used during iterations when alternating least squares is applied and the maximum variance of the raw measurement data is explained. The profiles in the bilinear model resolved by MCR are physically and chemically meaningful and correspond to interpretable patterns of variation of principal components [27, 54, 55].

In our study, MCR-ALS was applied to resolve binary amino acid mixtures of L-glutamic acid and L-glutamine analyzed by THz-TDS. Non-negativity constraint was applied on both spectra and concentration directions during the iterations. The spectra of the pure analytes obtained from the MCR displayed are corresponding to glutamine and glutamic acid. MCR results provided fitting error in % (exp) is 6.731, and percent of variance explained (R^2) at the optimum is 99.55, and this results show that the MCR model can fit the raw data well in this case. The correlation coefficients (r^2) between the reference THz spectra of pure analytes and those resolved pure spectra for each principal component by MCR are 0.9990 and 0.9979 for glutamine and glutamic acid, respectively. The spectrum of glutamine resolved by MCR is in good agreement with that measured by THz in the laboratory. The fitting constant is 0.9999.

MCR-ALS was also applied to resolve the binary isomer mixtures of D-(−)fructose and D-(+) galactose analyzed by THz-TDS. The absorption spectra of the two components obtained by MCR-ALS were well fitting to spectra of pure D-(−)fructose and D-(+)galactose anhydrous, which were obtained from experiment data in the same condition, respectively. The results of correlation coefficient between THz spectra obtained by MCR-ALS and pure D-(−)fructose and D-(+)galactose samples are 0.9974 and 0.9933, respectively. Relative concentrations of the two components were resolved by MCR-ALS, and they can fit the true concentrations well. MCR-ALS successfully resolved pure THz spectra of the components in the binary isomer mixtures and their corresponding concentrations.

Hyperspectral images based on infrared, near-infrared, Raman, and fluorescence are an active area of research that has grown quickly since a decade ago [56]. THz time-domain imaging is an emerging modality and has attracted a lot of interest since THz spectra got successes [57]. MCR-ALS can easily resolve the pure spectra and their corresponding concentration distributions from hyperspectral imaging, as well as external spectral can be used for local rank constraints. In the future, MCR-ALS can be proposed to resolve the biomedical images based on THz spectroscopy based on the development of THz hyperspectral imaging [56]. We believe that THz time-domain imaging will provide potential for medical diagnosis.

4. Does terahertz radiation lead to DNA or tissue damage?

Investigations of the interaction between nonionizing electromagnetic radiation and biological systems are necessary before the application of electromagnetic radiation in medical diagnosis. Recent emergence and growing use of terahertz radiation for medical imaging and public security screening raise questions on reasonable levels of exposure and health consequences of it. In particular, picosecond-duration THz pulses have shown promise for novel diagnostic imaging techniques. From different studies, the researcher got different conclusions on the effects of THz pulses on human cells and tissues. Titova et al. studied the biological effects of THz radiation on artificial human skin tissues [58], and their work shows that THz pulse irradiation may cause DNA damage in exposed skin tissue when intense THz pulses are applied. They consider that DNA damage repair mechanisms are quickly activated after THz radiation. But they found that the cellular response to pulsed THz radiation is significantly different from that induced by exposure to UVA (400 nm). However, Hintzsche et al. investigated power intensities ranged from 0.03 to 0.9 mW/cm and the cells were exposed for 2 and 8 h. Chromosomal damage and DNA damage were not detected in the disdained condition. Cell proliferation was also found to be unaffected by the exposure [59]. Bogomazova studied DNA damage and transcriptome responses in human embryonic stem cells (hESCs). They did not observe any effect on the mitotic index or morphology of the hESCs following THz exposure [60]. Anyway, THz is still one of the most harmless techniques for biomolecular analysis and medical diagnosis, and it will be powerful to detect the information from DNA, protein, or tissues.

5. Conclusion

Terahertz technology is progressing in biological and medical diagnosis in recent years. It has potential to be applied on analysis of biological molecular, cellular, tissues and organs, since terahertz is a nondestructive technique. The study displayed in this chapter indicates that terahertz spectroscopy and terahertz imaging technology combined with chemometrics can give accurate classification for normal and cancer tissues, predict the concentration of different compounds, and resolve the pure THz spectra of biomolecules in mixture systems. Comparing the classification accuracies pretreated by different preprocessing methods,

it indicated that the models applied based on terahertz spectroscopy could provide a better application for medical diagnosis. Terahertz is still one of the most harmless techniques for medical diagnosis, and it is hopefully to be developed as a powerful technique to detect the information from biomolecules or tissues in the future.

Acknowledgements

Author details

Xin Zhang and Zhuoyong Zhang*

*Address all correspondence to: zhangzhuoyong@cnu.edu.cn

Department of Chemistry, Capital Normal University, Beijing, China

References

[1] B. Yu, F. Zeng, Y. Yang, Q. Xing, A. Chechin, X. Xin, I. Zeylikovich, and R. R. Alfano, "Torsional vibrational modes of tryptophan studied by terahertz time-domain spectroscopy," *Biophysical Journal*, vol. 86, pp. 1649–1654, Mar 2004.

[2] M. Walther, P. Plochocka, B. Fischer, H. Helm, and P. U. Jepsen, "Collective vibrational modes in biological molecules investigated by terahertz time-domain spectroscopy," *Biopolymers*, vol. 67, pp. 310–313, 2002.

[3] A. J. Fitzgerald, E. Berry, N. N. Zinovev, G. C. Walker, M. A. Smith, and J. M. Chamberlain, "An introduction to medical imaging with coherent terahertz frequency radiation," *Physics in Medicine and Biology*, vol. 47, pp. R67–R84, Apr 7 2002.

[4] N. Qi, Z. Zhang, Y. Xiang, Y. Yang, and P. d. B. Harrington, "Terahertz time-domain spectroscopy combined with fuzzy rule-building expert system and fuzzy optimal associative memory applied to diagnosis of cervical carcinoma," *Medical Oncology*, vol. 32, pp. 1–6, 2015.

[5] M. Brucherseifer, M. Nagel, P. H. Bolivar, H. Kurz, A. Bosserhoff, and R. Büttner, "Label-free probing of the binding state of DNA by time-domain terahertz sensing," *Applied Physics Letters*, vol. 77, pp. 4049–4051, 2000.

[6] J. Xu, K. W. Plaxco, and S. J. Allen, "Probing the collective vibrational dynamics of a protein in liquid water by terahertz absorption spectroscopy," *Protein Science*, vol. 15, pp. 1175–1181, 2006.

[7] H. Yoshida, Y. Ogawa, Y. Kawai, S. Hayashi, A. Hayashi, C. Otani, E. Kato, F. Miyamaru, and K. Kawase, "Terahertz sensing method for protein detection using a thin metallic mesh," *Applied Physics Letters*, vol. 91, p. 253901, 2007.

[8] S. Nakajima, H. Hoshina, M. Yamashita, C. Otani, and N. Miyoshi, "Terahertz imaging diagnostics of cancer tissues with a chemometrics technique," *Applied Physics Letters*, vol. 90, p. 041102, 2007.

[9] G. Sozzi, D. Conte, L. Mariani, S. L. Vullo, L. Roz, C. Lombardo, M. A. Pierotti, and L. Tavecchio, "Analysis of circulating tumor DNA in plasma at diagnosis and during follow-up of lung cancer patients," *Cancer Research*, vol. 61, pp. 4675–4678, 2001.

[10] M. Sirajuddin, S. Ali, and A. Badshah, "Drug–DNA interactions and their study by UV–Visible, fluorescence spectroscopies and cyclic voltametry," *Journal of Photochemistry and Photobiology, B: Biology*, vol. 124, pp. 1–19, 2013.

[11] D. L. Sokol, X. Zhang, P. Lu, and A. M. Gewirtz, "Real time detection of DNA RNA hybridization in living cells," *Proceedings of the National Academy of Sciences*, vol. 95, pp. 11538–11543, 1998.

[12] J. Khandurina, T. E. McKnight, S. C. Jacobson, L. C. Waters, R. S. Foote, and J. M. Ramsey, "Integrated system for rapid PCR-based DNA analysis in microfluidic devices," *Analytical Chemistry*, vol. 72, pp. 2995–3000, 2000.

[13] S. Stranieri, M. Karna, and D. Shreiber, "Terahertz Characterization of DNA: Enabling a Novel Approach". No. ARL-CR-0788. ILLINOIS UNIV AT URBANA-CHAMPAIGN, p.20, 2015.

[14] M. Nagel, P. H. Bolivar, M. Brucherseifer, H. Kurz, A. Bosserhoff, and R. Büttner, "Integrated THz technology for label-free genetic diagnostics," *Applied Physics Letters*, vol. 80, pp. 154–156, 2002.

[15] D. Polley, A. Patra, and R. K. Mitra, "Dielectric relaxation of the extended hydration sheathe of DNA in the THz frequency region," *Chemical Physics Letters*, vol. 586, pp. 143–147, 2013.

[16] K. A. Niessen, M. Xu, E. Snell, and A. Markelz, "THz microscopy measurements on inhibitor dependence of protein intramolecular modes," in *2013 38th International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz)*, Mainz, Germany, 1–6 September 2013.

[17] T. Chen, Z. Li, and M. Wei, "Identification of biomolecules by terahertz spectroscopy and fuzzy pattern recognition," *Spectrochimica Acta Part A Molecular & Biomolecular Spectroscopy*, vol. 106, pp. 48–53, 2013.

[18] N. Qi, Z. Zhang, and Y. Xiang, "Application of terahertz technology in medical testing and diagnosis," *Spectroscopy and Spectral Analysis*, vol. 33, pp. 2064–2070, Aug 2013.

[19] E. Jung, M. Lim, K. Moon, Y. Do, S. Lee, H. Han, H. J. Choi, K. S. Cho, and K. R. Kim, "Terahertz pulse imaging of micro-metastatic lymph nodes in early-stage cervical cancer patients," *Journal of the Optical Society of Korea*, vol. 15, pp. 155–160, Jun 2011.

[20]   S. Smye, J. Chamberlain, A. Fitzgerald, and E. Berry, "The interaction between terahertz radiation and biological tissue," *Physics in Medicine and Biology*, vol. 46, p. R101, 2001.

[21]   R. M. Woodward, B. E. Cole, V. P. Wallace, R. J. Pye, D. D. Arnone, E. H. Linfield, and M. Pepper, "Terahertz pulse imaging in reflection geometry of human skin cancer and skin tissue," *Physics in Medicine and Biology*, vol. 47, p. 3853, 2002.

[22]   M. M. Nazarov, A. P. Shkurinov, E. A. Kuleshov, and V. V. Tuchin, "Terahertz time-domain spectroscopy of biological tissues," *Quantum Electronics*, vol. 38, p. 647, 2008.

[23]   R. M. Woodward, B. E. Cole, V. P. Wallace, R. J. Pye, D. D. Arnone, E. H. Linfield, and M. Pepper, "Terahertz pulse imaging in reflection geometry of human skin cancer and skin tissue," *Physics in Medicine and Biology*, vol. 47, pp. 3853–3863, Nov 7 2002.

[24]   P. Knobloch, C. Schildknecht, T. Kleine-Ostmann, M. Koch, S. Hoffmann, M. Hofmann, E. Rehberg, M. Sperling, K. Donhuijsen, G. Hein, and K. Pierz, "Medical THz imaging: an investigation of histo-pathological samples," *Physics in Medicine and Biology*, vol. 47, pp. 3875–3884, Nov 7 2002.

[25]   O. P. Cherkasova, M. M. Nazarov, E. E. Berlovskaya, A. A. Angeluts, A. M. Makurenkov, and A. P. Shkurinov, "Studying human and animal skin optical properties by terahertz time-domain spectroscopy," *Bulletin of the Russian Academy of Sciences: Physics*, vol. 80, pp. 479–483, 2016.

[26]   M. A. Sharaf, D. L. Illman, and B. R. Kowalski, *Chemometrics* vol. 82: John Wiley & Sons: New York, 1986.

[27]   S. Lu, X. Zhang, Z. Zhang, Y. Yang, and Y. Xiang, "Quantitative measurements of binary amino acids mixtures in yellow foxtail millet by terahertz time domain spectroscopy," *Food Chemistry*, vol. 211, pp. 494–501, 2016.

[28]   J. El Haddad, B. Bousquet, L. Canioni, and P. Mounaix, "Review in terahertz spectral analysis," *TrAC, Trends in Analytical Chemistry*, vol. 44, pp. 98–105, 2013.

[29]   M. Otsuka, J. i. Nishizawa, J. Shibata, and M. Ito, "Quantitative evaluation of mefenamic acid polymorphs by terahertz–chemometrics," *Journal of Pharmaceutical Sciences*, vol. 99, pp. 4048–4053, 2010.

[30]   Y. Hua, H. Zhang, and H. Zhou, "Quantitative determination of cyfluthrin in n-hexane by terahertz time-domain spectroscopy with chemometrics methods," in *Instrumentation and Measurement Technology Conference, 2009. I2MTC'09. IEEE*, 2009, pp. 670–674.

[31]   J. El Haddad, F. de Miollis, J. Bou Sleiman, L. Canioni, P. Mounaix, and B. Bousquet, "Chemometrics applied to quantitative analysis of ternary mixtures by terahertz spectroscopy," *Analytical Chemistry*, vol. 86, pp. 4927–4933, 2014.

[32]   F. Ellrich, G. Torosyan, S. Wohnsiedler, S. Bachtler, A. Hachimi, J. Jonuscheit, R. Beigang, F. Platte, K. Nalpantidis, and T. Sprenger, "Chemometric tools for analysing terahertz fingerprints in a postscanner," in *2012 37th International Conference on Infrared, Millimeter, and Terahertz Waves*, Wollongong, NSW, Australia, 23–28 September 2012.

[33] N. Qi, Z. Zhang, Y. Xiang, Y. Yang, X. Liang, and P. d. B. Harrington, "Terahertz time-domain spectroscopy combined with support vector machines and partial least squares-discriminant analysis applied for the diagnosis of cervical carcinoma," *Analytical Methods*, vol. 7, pp. 2333–2338, 2015.

[34] A. Savitzky and M. J. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Analytical Chemistry*, vol. 36, pp. 1627–1639, 1964.

[35] H. H. Madden, "Comments on the Savitzky-Golay convolution method for least-squares-fit smoothing and differentiation of digital data," *Analytical Chemistry*, vol. 50, pp. 1383–1386, 1978.

[36] A. X. Zhao, X. J. Tang, Z. H. Zhang, and J. H. Liu, "The parameters optimization selection of Savitzky-Golay filter and its application in smoothing pretreatment for FTIR spectra," in *2014 9th IEEE Conference on Industrial Electronics and Applications*, Hangzhou, China, 9–11 June 2014, pp. 516–521.

[37] T. Isaksson and T. Næs, "The effect of multiplicative scatter correction (MSC) and linearity improvement in NIR spectroscopy," *Applied Spectroscopy*, vol. 42, pp. 1273–1284, 1988.

[38] M. Maleki, A. Mouazen, H. Ramon, and J. De Baerdemaeker, "Multiplicative scatter correction during on-line measurement with near infrared spectroscopy," *Biosystems Engineering*, vol. 96, pp. 427–433, 2007.

[39] T. Fearn, "On orthogonal signal correction," *Chemometrics and Intelligent Laboratory Systems*, vol. 50, pp. 47–52, 2000.

[40] J. Zhang, Z. Zhang, Y. Xiang, Y. Dai, and P. d. B. Harrington, "An emphatic orthogonal signal correction-support vector machine method for the classification of tissue sections of endometrial carcinoma by near infrared spectroscopy," *Talanta*, vol. 83, pp. 1401–1409, 2011.

[41] Y. Roggo, P. Chalus, L. Maurer, C. Lema-Martinez, A. Edmond, and N. Jent, "A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 44, pp. 683–700, 2007.

[42] Z. M. Zhang, S. Chen, and Y. Z. Liang, "Baseline correction using adaptive iteratively reweighted penalized least squares," *Analyst*, vol. 135, pp. 1138–1146, 2010.

[43] P. H. Eilers, "A perfect smoother," *Analytical Chemistry*, vol. 75, pp. 3631–3636, 2003.

[44] X. Zhang, A. de Juan, and R. Tauler, "Multivariate curve resolution applied to hyperspectral imaging analysis of chocolate samples," *Applied Spectroscopy*, vol. 69, pp. 993–1003, 2015.

[45] C. Torrence and G. P. Compo, "A practical guide to wavelet analysis," *Bulletin of the American Meteorological Society*, vol. 79, pp. 61–78, 1998.

[46] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 2, pp. 37–52, 1987.

[47] P. Geladi and B. R. Kowalski, "Partial least-squares regression: a tutorial," *Analytica Chimica Acta*, vol. 185, pp. 1–17, 1986.

[48] M. Barker and W. Rayens, "Partial least squares for discrimination," *Journal of Chemometrics*, vol. 17, pp. 166–173, 2003.

[49] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the Fifth Annual Workshop on Computational Learning theory*, Pittsburgh, Pennsylvania, USA. 1992, pp. 144–152.

[50] C. C. Chang and C. J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology* (*TIST*), vol. 2, p. 27, 2011.

[51] P. d. B. Harrington, "Fuzzy multivariate rule-building expert systems: minimal neural networks," *Journal of Chemometrics*, vol. 5, pp. 467–486, 1991.

[52] B. W. Wabuyele and P. d. B. Harrington, "Fuzzy optimal associative memory for background prediction of near-infrared spectra," *Applied Spectroscopy*, vol. 50, pp. 35–42, 1996.

[53] P. d. B. Harrington, "Statistical validation of classification and calibration models using bootstrapped Latin partitions," *TrAC Trends in Analytical Chemistry*, vol. 25, pp. 1112–1124, 2006.

[54] S. Navea, A. d. Juan, and R. Tauler, "Detection and resolution of intermediate species in protein folding processes using fluorescence and circular dichroism spectroscopies and multivariate curve resolution," *Analytical Chemistry*, vol. 74, pp. 6031–6039, 2002.

[55] E. Pere-Trepat, M. Petrovic, D. Barcelo, and R. Tauler, "Application of chemometric methods to the investigation of main microcontaminant sources of endocrine disruptors in coastal and harbour waters and sediments," *Analytical and Bioanalytical Chemistry*, vol. 378, pp. 642–654, Feb 2004.

[56] S. Piqueras, L. Duponchel, R. Tauler, and A. De Juan, "Resolution and segmentation of hyperspectral biomedical images by multivariate curve resolution-alternating least squares," *Analytica Chimica Acta*, vol. 705, pp. 182–192, 2011.

[57] Z. Xu and E. Y. Lam, "Image reconstruction using spectroscopic and hyperspectral information for compressive terahertz imaging," *Journal of the Optical Society of America A: Optics and Image Science, and Vision*, vol. 27, pp. 1638–1646, 2010.

[58] L. V. Titova, A. K. Ayesheshim, A. Golubov, D. Fogen, R. Rodriguez-Juarez, F. A. Hegmann, and O. Kovalchuk, "Intense THz pulses cause H2AX phosphorylation and activate DNA damage response in human skin tissue," *Biomedical Optics Express*, vol. 4, pp. 559–568, 2013/04/01 2013.

[59] H. Hintzsche, C. Jastrow, B. Heinen, K. Baaske, T. Kleine-Ostmann, M. Schwerdtfeger, M. K. Shakfa, U. Kärst, M. Koch, and T. Schrader, "Terahertz radiation at 0.380 THz and 2.520 THz does not lead to DNA damage in skin cells in vitro," *Radiation Research*, vol. 179, pp. 38–45, 2012.

[60] A. Bogomazova, E. Vassina, T. Goryachkovskaya, V. Popik, A. Sokolov, N. Kolchanov, M. Lagarkova, S. Kiselev, and S. Peltek, "No DNA damage response and negligible genome-wide transcriptional changes in human embryonic stem cells exposed to terahertz radiation," *Scientific Reports*, vol. 5, p. 7749, 2015.