Computers and Chemical Engineering 143 (2020) 107099

Contents lists available at ScienceDirect

Computers and Chemical Engineering

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

Opportunities in tensorial data analytics for chemical and biological manufacturing processes

Weike Sun, Richard D. Braatz*

Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

ARTICLE INFO

Article history:
Received 2 May 2020
Revised 30 June 2020
Accepted 12 September 2020
Available online 24 September 2020

Keywords:
Process data analytics
Tensorial data analytics
Multilinear subspace learning
Tensor decomposition
Chemical process systems

ABSTRACT

With the development of technology in data collection and storage, new types of higher order tensorial information streams are available in chemical and biological manufacturing processes, which contain valuable information about the process condition and product quality. However, tensorial data have not been fully utilized yet and the application of tensorial data analytics to manufacturing processes has not been thoroughly investigated. In this article, different types of higher order data in manufacturing processes are described, and their potential usage is addressed. Then some perspectives are provided on the application of tensorial data analytics to manufacturing processes, with an emphasis on multilinear subspace learning problems. In particular, the most representative multilinear subspace learning methods are reviewed. Looking into the future, the potential and research needs for tensorial data analytics are briefly discussed.

Contents

1. Introduction

Process data analytics, which refer to the application of machine learning techniques to manufacturing data, are becoming increasingly popular due to its capability of improving process productivity, reliability, and control. With the development in sensor

* Corresponding author. 77 Massachusetts Avenue, Room E19-551, Cambridge, MA 02139.
E-mail address: braatz@mit.edu (R.D. Braatz).

https://doi.org/10.1016/j.compchemeng.2020.107099
0098-1354/

technologies and wireless networks, an increase in the availability of various types of data and a decrease in costs for data collection and storage further enable wide application of process data analytics. Process data analytics have been applied in various chemical and biological manufacturing processes, from the level of individual unit operations up to the level of the entire manufacturing systems. It can be used in a number of different ways by the manufacturers. For example, the process models can be used to make predictions, such as the early identification of batches which will eventually fail to meet specifications, or to modify operating procedures for downstream processes to improve product quality (Chiang et al., 2017; Mitchell, 2014). Process data analytics can be used for control, that is, to compute adjustments to the critical process parameters to move the quality variables towards desirable values (Moe et al., 2018), and can also be used for the real-time detection and diagnosis of anomalous behavior, which is referred to as *process monitoring* and *statistical process control* in the chemical engineering literature (Chiang et al., 2000; Ge et al., 2013; Qin, 2012; Severson et al., 2016).

Due to the high dimensionality of manufacturing data and the volume of measurements, linear subspace learning has been widely applied in the chemical and biotech industries for dimensionality reduction, pattern recognition, and data mining. For example, principal component analysis (PCA) (Jackson, 1991) is an explanatory tool to handle high-dimensional data by extracting the major source of variation from highly correlated measurements, and is widely used for data visualization, information extraction, and process monitoring (MacGregor and Kourti, 1995; Wang et al., 2015). Popular supervised subspace learning methods for predictive modeling of high-dimensional data are principal component regression (PCR) and partial least squares (PLS) (Geladi and Kowalski, 1986; MacGregor, 2004; Næs and Martens, 1988). PCR and PLS construct regression models using the derived latent variables, which not only improve the modeling accuracy for high-dimensional noisy measurements, but the derived latent variables can also be useful for process understanding.

The widely applied process analytical methods are designed to be applied to a data matrix (aka two-way array aka second-order tensor), such as a collection of vector measurements or a single gray-scale picture. With the increasing usage of multisensory technology and data storage infrastructure, new types of higher order multidimensional data arrays (aka tensors) arise in a variety of chemical and biological manufacturing processes. Instead of vector measurements, a single measurement can consist of second-, third-, or higher order tensors. One example is the coupling of on-line automated sampling systems with liquid chromatography-mass spectroscopy (LC-MS) to simultaneously measure transient variations in large numbers of distinct small molecules and proteins during pharmaceutical manufacturing (Gülbakan et al., 2015; Rogers et al., 2015; Jiang et al., 2017). As a standard practice in the industry, the higher order data are often pre-processed and reformatted as large matrices and then the classical two-way analysis methods are applied to the pre-processed data (Bro, 1999; Jeffy et al., 2018). However, this approach is not optimal, and important structural information within multiway data cannot be discovered. To this end, tensorial data analytics methods have been developed to directly address higher order data for process monitoring, design, and optimization, which has not gained enough attention in the chemical engineering field and is what motivates this review article.

This article does not review the entire tensorial data analytics field or provide detailed discussions on all of the various tensorial data analytics techniques. The focus of this article is to provide some perspectives on the potential usage of tensorial data in chemical and biological systems and address the value of the direct application of tensorial data analytics. This review of tensorial data analytics emphasizes multilinear subspace learning, as this approach is the largest class of methods and is the direct extension of the matrix methods that are widely applied to chemical and biological systems. The article then discusses some future research needs.

2. New information streams: tensorial data

Modern datasets collected in the chemical and biological manufacturing processes not only have scalars (e.g., temperature measurement) and 1-way arrays (e.g., multiple sensors or a single spectrum) as a single measurement, but also 2-way arrays (e.g., measurements from batch processes or gray-scale images), 3-way arrays (e.g., color or hyperspectral images), and even higher order arrays (e.g., color videos) . Here the *order* of the data refers to a single sample. When analyzing multiple measurements at the same time, which is usually the case for process data analytics, the order of the dataset is increased by one representing the sample number. This section describes some examples of various higher order datasets in manufacturing processes and their usage for process design and optimization.

2.1. Examples of two-way arrays

A typical example of a single two-way array measurement is a matrix of measurements collected from a single batch process run. Each batch sample is a series of measurements collected over a period of time on a separate, identifiable item or parcel of material. Such batch processes are common in fine chemicals and (bio)pharmaceutical manufacturing, for both the active pharmaceutical ingredient and the drug product. Each batch measurement is taken over a specific run with the first order representing the sample number and the second order representing different sensor measurements. Batch data can be used to design a model for early prediction of failure of a batch or use the data to build a model using critical process parameters to predict critical quality attributes. For example, a real-time monitoring system is developed to accurately identify the end-point of the batch in order to reduce the overall cycle time of the process as described in (Marjanovic et al., 2006).

Another type of two-way measurement is the gray-scale image. The images captured during the process can be used as an efficient non-invasive low-cost analysis for process monitoring and product quality control. An example is the gray-scale stereomicroscope images of product crystals from a slug-flow crystallizer (see Fig. 1). Each image is a 2-way matrix with each order representing space along horizontal and vertical axes and each pixel being a gray-scale value between 0 and 255. The product crystal size and shape statistics can be directly obtained by analyzing the images. Besides, the performance of the designed slug-flow crystallizer, a relatively low degree of aggregation with high slug-to-slug variability as shown in Fig. 1, can be obtained from the figure to infer appropriate adjustment of operating condition.

The last example is the measurement collected by hyphenated techniques (Patel et al., 2010). One typical hyphenated technique is LC-MS. The measured data has two orders with the first order representing retention time from LC and the second order representing mass-to-charge ratio from MS, as illustrated in Fig. 2. The traditionally used high-performance liquid chromatography (HPLC) method lacks specificity for some molecular mixtures. LC-MS combines the separation capabilities of HPLC and the mass analysis capabilities of mass spectrometry to provide very sensitive and specific detection of a wide range of molecules. Decreased price and easier implementation has enabled the application of LC-MS in manufacturing processes. The coupling of on-line automated sampling systems with LC-MS simultaneously

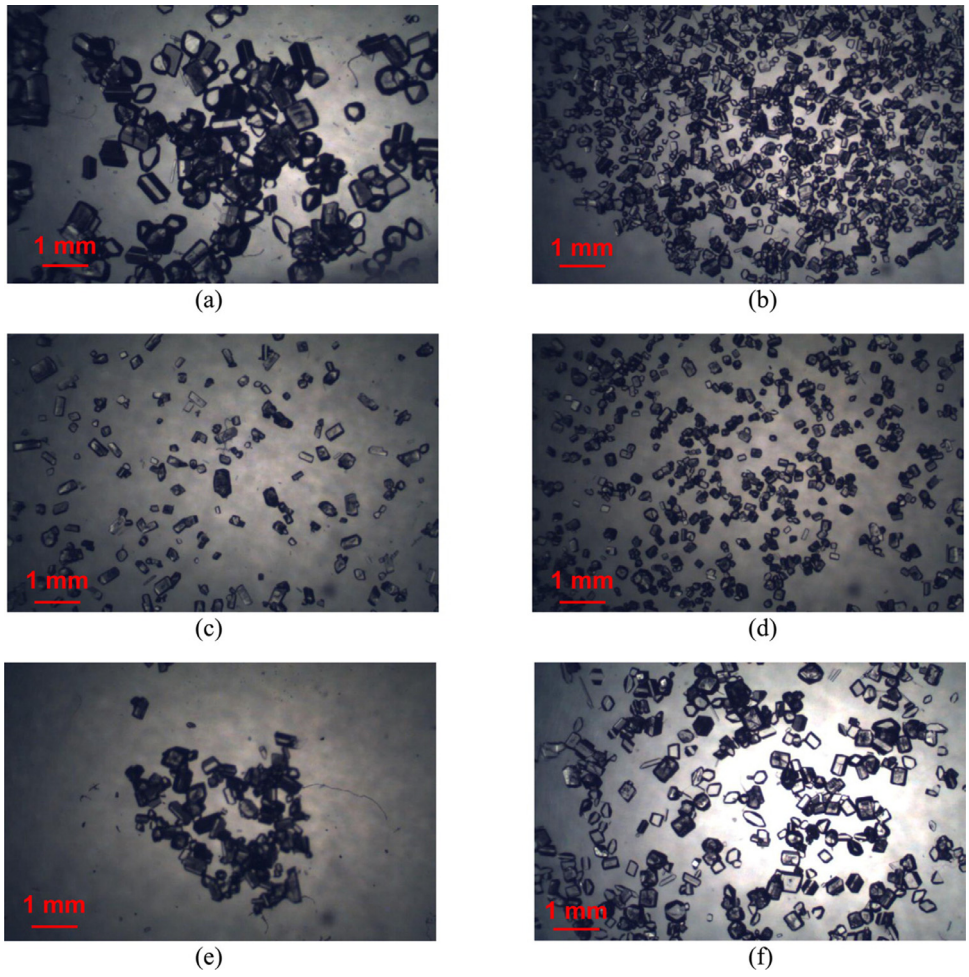

Fig. 1. gray-scale stereomicroscope images of product crystals with nucleation induced by coaxial mixing in different experimental runs, adapted from (Jiang et al., 2014).

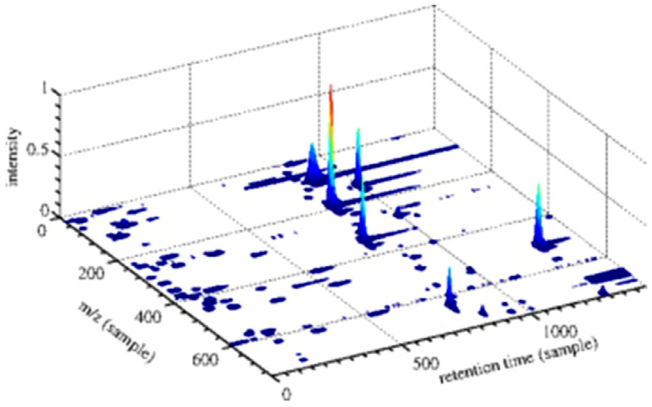

Fig. 2. Illustration of an LC-MS measurement, adapted from (Rapin et al., 2016).

measures transient variations in large numbers of distinct small molecules and proteins during biomanufacturing, which is useful for product assessment and process monitoring. Additional types of analytical chemistry techniques that provide matrix measurements include liquid chromatography with photodiode array detection (LC-DAD), gas chromatography-mass spectrometry (GC–MS), fluorescence, and MS/MS (Seger et al., 2005; Hübschmann, 2015; Jiang et al., 2017; Ou-Yang et al., 2018).

2.2. Examples of three-way arrays

The simplest type of a three-way array measurement is coupling two-way arrays with time. For example, a gray-scale video consists of gray-scale images over time, where the first two orders are for the x- and y-axis of images and the third order is for time. Similarly, when taking LC-MS over a period of time, the extra order is the time point.

Another example of a three-way measurement is spectral imaging. An ordinary camera captures color images with the order of the data increases by one as compared to the gray-scale image. The additional order being the color axis for red, green, and blue (RGB). The data are stored as a number between 0 and 255 for red, green, and blue at each pixel of a two-way array. Color images have been used in industry to assess process performance. For example, the RGB image captured during the flotation process is utilized for monitoring and control of the process by analyzing the bubble size distribution, the presence and amount of clear windows, or black holes in the froth (Liu et al., 2005).

Besides traditional RGB imaging, spectral imaging encompasses a wide variety of techniques, such as infrared, visible spectrum, and x-rays. Hyperspectral imaging is a subcategory of spectral imaging, which collects a continuous spectral band at every pixel in an image plane. One application of hyperspectral imaging is the near-infrared chemical imaging (NIR-CI) for process monitoring of a freeze-drying process (Brouckaert et al., 2018). Freeze drying is a method for preserving the chemical nature of substances with sensitive thermal reactivity, which is useful for biological drug

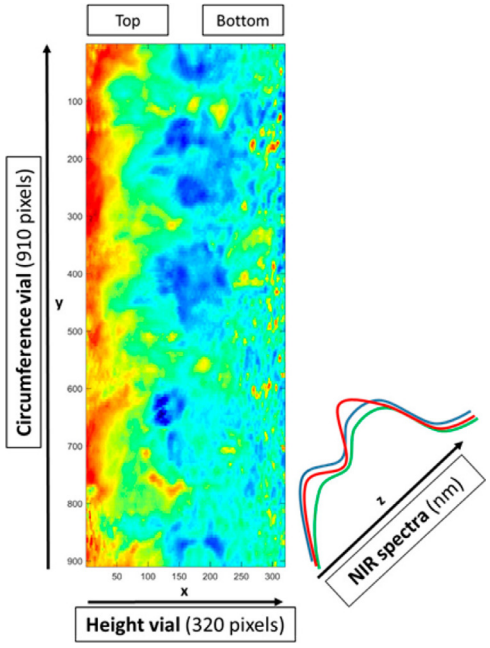

Fig. 3. NIR-CI of a freeze-drying sample. The color scale shows the highest pixels in red and lowest in blue, adapted from (Brouckaert et al., 2018).

storage. Freeze-drying processes have had very limited in-process analytical sensor technology, and a potential solution is to integrate imaging-based process analytical technology devices to monitor operations and ensure final product quality. The NIR-CI captured during the freeze-drying process (see Fig. 3) has a three-way structure, where the x and y axes are the height and full circumference of the freeze-dried cake, and a NIR spectrum is obtained for each pixel in the image as shown in the z-axis. The NIR-CI combines the chemical selectivity of spectroscopy with spatial information, which can be used for water content determination as in traditional NIR spectroscopy. NIR-CI also opens up possibilities for assessing the homogeneity throughout the product via individual detector elements in a CI array. Another example is the IR thermography of a freeze-drying process (Emteborg et al., 2014), where an IR camera is shown to be a highly versatile tool for online monitoring (see Fig. 4). An IR camera can take images every 120 s and provides contact-free measurements of the temperature distribution of the freeze-drying shelf, which offers superior spatial and thermal resolution in contrast to traditional probes. The application of IR camera for process monitoring should allow better design of shelves and trays in the freeze dryer, and could potentially provide a better understanding of how different materials are dried during different steps.

Samples collected from several hyphenated chromatographic and multidimensional chromate-graphy techniques also have a third-order structure. For example, the measurement from two-dimensional gas chromatography-mass spectrometry (GC × GC–MS) is a third-order tensor. The GC × GC–MS uses two GC columns for separation where the second column is operated with a different stationary phase. The third order represents the mass-to-charge ratio from MS. GC × GC–MS has been applied to the analytical characterization of complex metabolomes (Koek et al., 2011; Winnike et al., 2015) and organic compounds (Ou-Yang et al., 2018), which gave more accurate chromatographic peak capacity, selectivity, and lower detection limit for the analysis of small molecules than GC–MS. GC × GC–MS is also a promising tool for large-scale broad-spectrum biomarker discovery, which requires new bioinformatics tools to process the data in an efficient and proper way (Shi et al., 2014).

2.3. Examples of four-way arrays

A typical example of a four-way array in the manufacturing process is the video. The first three orders represent a single image while the last order represents the time. One application example is the inline imaging system used to characterize the shape properties of crystals in liquid slugs that flow down a tubular crystallizer (see Fig. 5). The low-cost imaging system is composed of a basic stereomicroscope and a video camera. Many such images as shown in Fig. 5 are collected each second in real-time video. This real-time video can be used to guide the experimental design, including the improvement of the slug aspect ratio, visualization of crystal shapes, and online monitoring of the extent of aggregation.

2.4. Remarks

Sections 2.1–2.3 describe several examples of tensorial data collected from the manufacturing process that could be used for better system design and operations. Many types of tensorial data will become increasingly common with the continued advances being made in sensor technology.

A particular fast-growing market is for hyperspectral imaging. According to a research report by Grand View Research (2019), the hyperspectral imaging system market size is estimated at $8.2 billion in 2017 and is expected to grow at 10.06% each year. Due to the increasing investment and technology development in hyperspectral imaging, now hyperspectral imaging systems can be acquired at a much lower price and are easier to use. In the past, a hyperspectral imaging system was expensive at about $50,000 to $100,000 for a single device, and required complex specialized hardware to operate. Hyperspectral imaging systems have become more affordable – with prices available from $1000 to $20,000 – and are easier to operate. In many cases, sample preparation is not required, and the optical systems are rugged and readily available. There are also compact designs for the hyperspectral imaging system, and results from specialized cell phones have been reported that take hyperspectral images directly (e.g., Kim et al., 2019; Salazar-Vazquez and Mendez-Vazquez, 2020). Those advances enable non-destructive acquisition of both qualitative and quantitative information from the manufacturing processes.

In summary, advances in technology are making available low-cost and user-friendly sensors for generating tensorial data. The tensorial information streams are expected to become widely used and open up new possibilities for more accurate and efficient process monitoring and design.

3. Tensorial data analytics

While tensorial data bring in more meaningful information for the establishment of high-fidelity models, the question becomes how to fully utilize such data. Due to the complexity of modern instruments, the volume of measurements taken, the high-dimensional nature, and the strong correlations of the manufacturing data, dimensionality reduction techniques are widely used for various analytical problems. For the traditional matrix-based data analytics, linear subspace learning methods have been widely applied in industry for both supervised and unsupervised learning problems. For unsupervised learning, such as data exploration and process monitoring, the typical methodologies include PCA and independent component analysis (Hyvärinen and Oja, 2000). For predictive modeling, PCR and PLS are powerful methods that account for multicollinearity of the predictors. For classification, such as

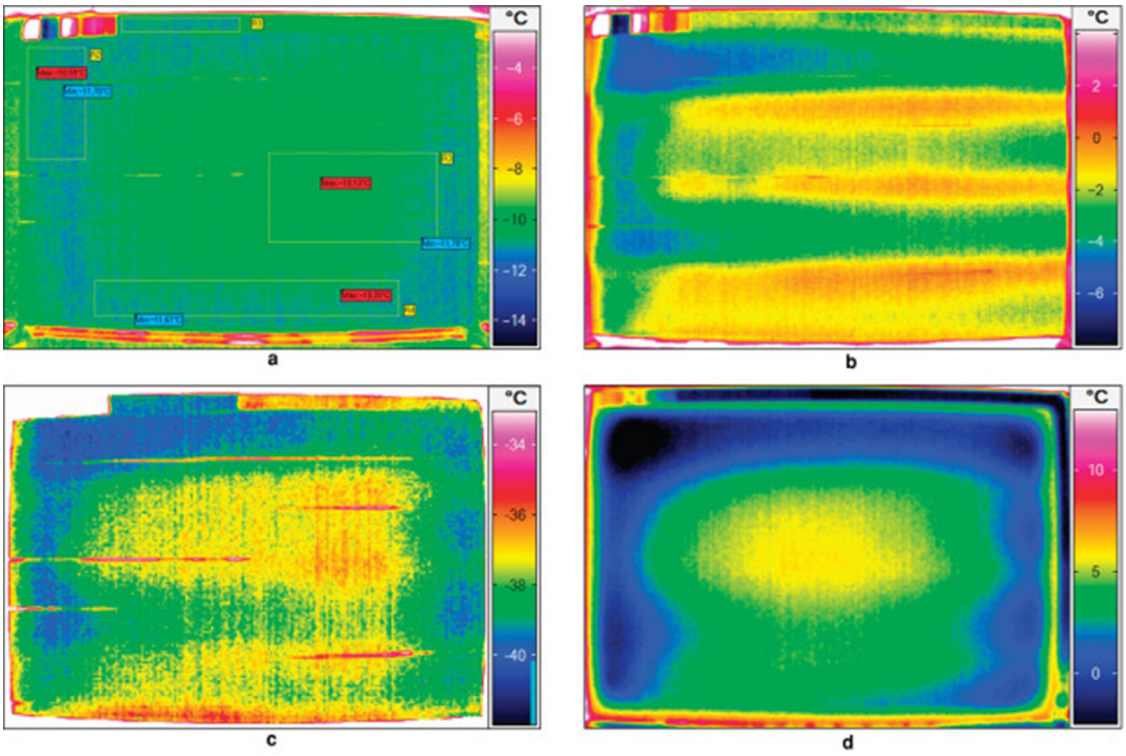

Fig. 4. IR images of the freeze-drying shelf with different running condition, adapted from (Emteborg et al., 2014).

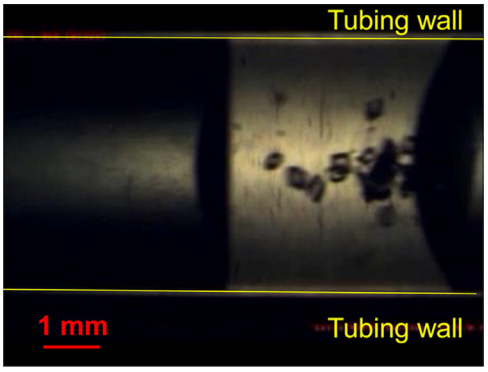

Fig. 5. An in-line stereomicroscope image from the real-time video of a crystallization process, adapted from (Jiang et al., 2014).

distinguishing bad batch production from good batches, Fisher discriminant analysis (Welling, 2009), aka linear discriminant analysis (LDA), is a widely applied linear dimensionality reduction technique for improving classification accuracy.

Traditionally, three approaches are widely used to pre-process tensorial data in order to use matrix-based methods for data analytics. The first approach is to apply a matrix-based method to each two-way dataset. For example, as shown in Fig. 6a, PCA is applied to each gray-scale image. This method cannot be applied to higher order tensors, and does not produce a single model for all of the measurements. The second approach is to unfold the tensor over specific orders to form a large matrix and then apply a matrix-based method. For example, as shown in Fig. 6b, the images are stacked to form a matrix, and a linear subspace learning method is applied to the new matrix. This approach has been used extensively in the chemical engineering community and is referred to as the *multiway method*. Multiway methods are the application of classic linear subspace learning methods to the unfolded data, which have been applied to a variety of processes (Hu and Yuan, 2008; J et al., 2018; Kourti, 2003; Lakshminarayanan et al., 1996; Lee et al., 2004; Marjanovic et al., 2006; Yu and Qin, 2009). The third approach, as shown in Fig. 6c, is to average the data over certain orders. For example, the average intensity is calculated over all images and PCA is performed on the averaged image.

The traditional methods are suboptimal as reformatting the data, and using only matrix-based methods potentially loses important structural information in the tensors. Besides, those methods do not provide the most compact way of data representation and often lead to overfitting. Finally, the constructed model is also difficult to interpret and might not improve process understanding. The ideal model should be constructed by unbiased data-driven approaches that directly extract information from the data tensor. Tensorial data analytics techniques are designed to handle tensors directly from its natural tensorial representation, and many different tensorial data analytics methods have been developed. In the rest of this section, the discussion will focus on multilinear subspace learning, which is a generalization of the linear subspace learning methods widely used in chemical engineering to tensors. The rest of this section discusses the extension of PCA to multilinear decomposition, as well as other methodologies, with the purpose of providing an introduction to the topic and the potential of these methods. Readers interested in a more exhaustive discussion of the methodologies including deep dives into the underlying mathematics are directed to existing reviews written for computer science and data science audiences (Cichocki, 2014; Cong et al., 2015; Lu et al., 2011; Papalexakis et al., 2016; Rabanser et al., 2017; Sidiropoulos et al., 2017).

3.1. From linear subspace learning to multilinear subspace learning

The generalization of linear subspace learning to multilinear subspace learning can be cast into a general framework, and the generalization of PCA to multilinear tensor decomposition is shown as an example here.

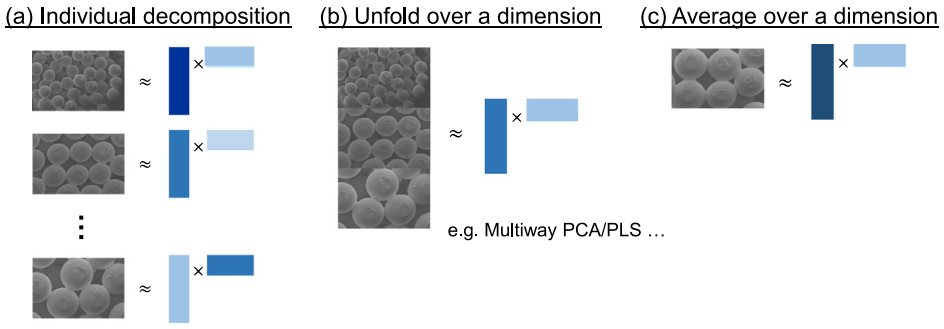

Fig. 6. Three traditional approaches to preprocess tensorial data to enable application of a matrix-based data analytics method (example images are from Borden et al., 2003).

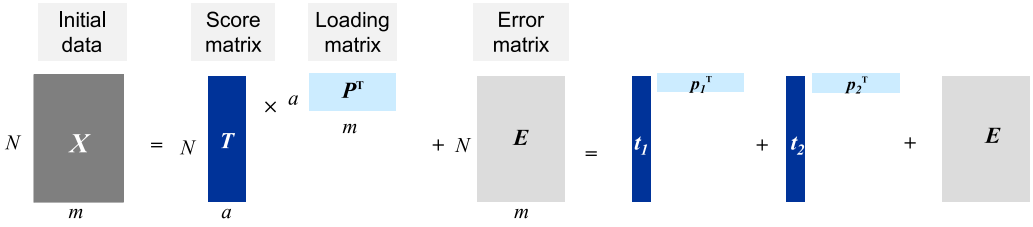

Fig. 7. Illustration of PCA based matrix decomposition.

Recall that PCA is a linear dimensionality reduction technique that is optimal in terms of maximizing the retained variability in the lower-dimensional representation. PCA calculates a set of orthogonal vectors, also called loading vectors, in an order of decreasing variance explained in the corresponding loading directions. Given a training set of N observations, m variables, and the corresponding training matrix $\boldsymbol{X} \in R^{N \times m}$, the loading vectors are calculated by solving the optimization

$$\max_{\boldsymbol{v} \neq 0} \frac{\boldsymbol{v}^\top \boldsymbol{X}^\top \boldsymbol{X} \boldsymbol{v}}{\boldsymbol{v}^\top \boldsymbol{v}} \tag{1}$$

where $\boldsymbol{v} \in R^m$ is the loading vector. The problem (1) can be solved via the singular value decomposition (SVD),

$$\frac{1}{\sqrt{N-1}} \boldsymbol{X} = \boldsymbol{U} \boldsymbol{\Sigma} \boldsymbol{V}^\top \tag{2}$$

where $\boldsymbol{U} \in R^{N \times N}$ and $\boldsymbol{V} \in R^{m \times m}$ are unitary matrices, $\boldsymbol{V}$ contains the loading vectors, and $\boldsymbol{\Sigma} \in R^{N \times m}$ contains the singular values of decreasing magnitude in its main diagonal and 0 in off-diagonal elements. Besides the formulation (1), PCA can also be viewed as minimizing the reconstruction error of the projection.

$$\boldsymbol{X} = \boldsymbol{T} \boldsymbol{P}^\top + \boldsymbol{E} \tag{3}$$

In order to use PCA for dimensionality reduction, the loading vectors corresponding to the largest a singular values are typically retained and stored in the loading matrix $\boldsymbol{P} \in R^{m \times a}$, and the original data matrix can be decomposed as a summation of the principal subspace and the residual subspace as shown in Eq. (3), where the principal subspace is the multiplication of the score matrix $\boldsymbol{T} \in R^{N \times a}$ and the loading matrix $\boldsymbol{P}$. Another representation of PCA-based dimensionality reduction is that the original data matrix is approximated by a sum of rank-one products using the score and loading vectors. These two representations are equivalent and are illustrated in Fig. 7.

Based on multilinear algebra, any method from linear subspace can be extended to multilinear subspace within the general formulation with two modifications: the multilinear projection method employed and the objective criterion to optimize. The first step is to replace the linear projection with a multilinear projection. There are different multilinear projection methods based on different forms of input and output, such as vector-to-vector projection, tensor-to-tensor projection, and tensor-to-vector projection (Lu et al., 2011). The second step is to reformulate the objective function for tensors. For example, in order to expand PCA into multilinear subspace learning, the projection of a single vector measurement can be replaced by n-mode matrix product of a tensor

$$\boldsymbol{t} = \boldsymbol{P}^\top \boldsymbol{x} \;\Rightarrow\; \boldsymbol{\Gamma} = \mathcal{A} \times_1 \boldsymbol{U}^1 \times_2 \cdots \times_k \boldsymbol{U}^k \tag{4}$$

where $\boldsymbol{x}$ is the measurement vector and $\boldsymbol{t}$ is the score vector in PCA. $\mathcal{A} \in R^{m_1 \times \cdots \times m_k}$ is a k^{th}-order tensor measurement, $\boldsymbol{\Gamma} \in R^{a_1 \times \cdots \times a_k}$ is a k^{th}-order core tensor with reduced dimension in each order, and $\boldsymbol{U}^i \in R^{m_i \times a_i}$ is the projection matrix for the i^{th}-order. Then the objective function is re-defined as minimizing the reconstruction error of the tensors,

$$\min \frac{1}{N} \sum_{i=1}^{N} \|\boldsymbol{x}_i - \hat{\boldsymbol{x}}_i\| \Rightarrow \min \frac{1}{N} \sum_{i=1}^{N} \|\mathcal{A}_i - \hat{\mathcal{A}}_i\|_F \tag{5}$$

where the distance between two tensors is measured by the Frobenius norm. Depending on the exact projection method, the objective function, and the constraints, there are different ways to extend PCA into multilinear subspace learning. The two most widely used multilinear dimension reduction methods are discussed in Sections 3.2 and 3.3, which can be viewed as a generalization of PCA into multilinear PCA.

3.2. CANDECOMP/PARAFAC (CP) decomposition

The first method is the *CP decomposition* (Carroll and Chang, 1970; Harshman, 1970; Kiers, 2000). Similar to PCA which can be expressed as a summation of rank-1 matrices, the CP decomposition approximates a tensor by a sum of component rank-one tensors. For example, given a third-order tensor $\mathcal{A} \in R^{m_1 \times m_2 \times m_3}$, the CP decomposition factorizes $\mathcal{A}$ as

$$\mathcal{A} = \sum_k \boldsymbol{a}_k \otimes \boldsymbol{b}_k \otimes \boldsymbol{c}_k + E = \sum_k \lambda_k \boldsymbol{u}_k^1 \otimes \boldsymbol{u}_k^2 \otimes \boldsymbol{u}_k^3 + E \tag{6}$$

where $\otimes$ is the Kronecker product of vectors; $\boldsymbol{a}_k$, $\boldsymbol{b}_k$, $\boldsymbol{c}_k$ for $k = 1, \ldots, K$ are vectors of dimension m_1, m_2, m_3 respectively; and $\boldsymbol{u}_k^1$,

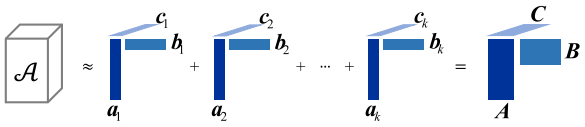

Fig. 8. CP decomposition of a three-way array.

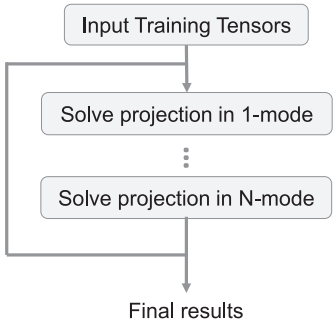

Fig. 9. General framework of the ALS algorithm to compute a CP decomposition.

$\boldsymbol{u}_k^2$, $\boldsymbol{u}_k^3$ are the normalized vectors of length one with weights absorbed into λ_k. The three-way decomposition is illustrated in Fig. 8. where ***A, B, C*** are the matrices with the columns being vectors $\boldsymbol{a}_k$, $\boldsymbol{b}_k$, $\boldsymbol{c}_k$, respectively. Note that the columns in matrices ***A, B, C*** could be linearly dependent, and is therefore different from PCA.

The CP decomposition can be applied to a general d^{th}-order tensor $\mathcal{A} \in R^{m_1 \times \cdots \times m_d}$, which is expressed as

$$\mathcal{A} = \sum_k \lambda_k \mathbf{u}_k^1 \otimes \mathbf{u}_k^2 \otimes \cdots \otimes \mathbf{u}_k^d + E \tag{7}$$

The objective function of the CP decomposition is to minimize the reconstruction error between the original tensor and the approximated tensor:

$$\min_{\lambda_k, \mathbf{u}_k^i} \left\| \mathcal{A} - \sum_{k=1}^{K} \lambda_k \mathbf{u}_k^1 \otimes \mathbf{u}_k^2 \otimes \cdots \otimes \mathbf{u}_k^d \right\|_F \tag{8}$$

Unlike PCA, the minimization (8) is not convex and does not have a closed-form solution, which requires a numerical optimization procedure. Given a fixed number of components K, one basic algorithm for the CP decomposition is alternating least squares (ALS) proposed in (Carroll and Chang, 1970; Harshman, 1970). The detailed procedure is omitted here and the reader can refer to the original paper. The key idea behind ALS is to fix all factor matrices except for one in order to optimize the non-fixed matrix and then repeat this step for each matrix until the stopping criterion is satisfied, as illustrated in Fig. 9.

An interesting property of the CP decomposition of higher order tensors is that the rank decompositions are often unique under certain mild conditions, whereas matrix decompositions are not. Uniqueness means that the CP decomposition provides the only possible combination of rank-one tensors that sums to $\mathcal{A}$, except for the elementary indeterminacies of scaling and permutation. As stated in (Kruskal, 1989, 1977), a sufficient condition for uniqueness of the CP decomposition of a third-order tensor is

$$k_A + k_B + k_C \geq 2K + 2 \tag{9}$$

where k_A denotes the k-rank of matrix ***A***. More discussion on the uniqueness of the CP decomposition can be referred to (Kolda and Bader, 2009). The uniqueness property is very useful in applications, which enables the exact determination of the underlying factors of what is being measured and is easy for interpretation, as discussed in (Bro, 1999; Williams et al., 2018).

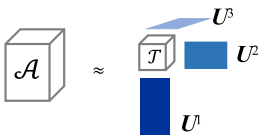

Fig. 10. Tucker decomposition of a three-way array.

3.3. Tucker decomposition

The second method is the *Tucker decomposition* (Tucker, 1966, 1963), which decomposes a tensor into a core tensor multiplied by a matrix along each mode, which can be viewed as a higher order PCA. For example, in the three-way case $\mathcal{A} \in R^{m_1 \times m_2 \times m_3}$

$$\hat{\mathcal{A}} = \mathcal{T} \times_1 \boldsymbol{U}^1 \times_2 \boldsymbol{U}^2 \times_3 \boldsymbol{U}^3 \tag{10}$$

where $\times_i$ is the i^{th} mode matrix product of a tensor, $\boldsymbol{U}^i \in R^{m_i \times t_i}$ are the factor matrices (which are usually orthogonal). The tensor $\mathcal{T} \in R^{t_1 \times t_2 \times t_3}$ is the core tensor and its entries show the level of interaction between different components. The Tucker decomposition of a three-way array is illustrated in Fig. 10.

Similar to the CP decomposition, the objective function for the Tucker decomposition of a d-way array is to minimize the reconstruction error

$$\min_{\mathcal{T},\ \boldsymbol{U}^i} \left\| \mathcal{A} - \mathcal{T} \times_1 \boldsymbol{U}^1 \times_2 \boldsymbol{U}^2 \cdots \times_d \boldsymbol{U}^d \right\|_F \tag{11}$$

A numerical optimization procedure is also required for solving Eq. (11). There are many algorithms for solving the Tucker decomposition, and one widely applied algorithm is called *higher order SVD* (HOSVD) (De Lathauwer et al., 2000), which imposes orthogonality constraints on the factor matrices and the core tensor is not super-diagonal. The key idea of HOSVD is to find the components that best capture the variation in each mode while fixing the other modes at that point. This directly corresponds to the basic PCA concept. The Tucker decomposition might be useful for tensor compression and exploratory analysis. A detailed discussion on different algorithms for the Tucker decomposition is available (Kolda and Bader, 2009). Unlike the CP decomposition, Tucker decompositions are generally not unique.

3.4. Available libraries

While the solution of tensorial data decomposition discussed in Sections 3.2 and 3.3 requires a numerical optimization procedure, many libraries in different programming languages are available to provide data structures for tensors and solutions of basic tensorial data analytics tools such as CP and Tucker decompositions. Here only a few popular tensor libraries are listed, which provide optimized approaches of storing tensors and algorithms for decomposing tensors. Libraries for Python, Matlab, and R are

- Tensor Toolbox for MATLAB [B. Bader et al., www.tensortoolbox.org]
- N-way Toolbox for MATLAB [R. Bro and C. Anderson, www.models.life.ku.dk/nwaytoolbox]
- Tensorlab for MATLAB [N. Vervliet et al., www.tensorlab.net/]
- Tensorly for Python [J. Kossaifi et al., tensorly.org/stable/home.html]
- TensorD for Python [M. Abadi et al., github.com/Large-Scale-Tensor-Decomposition/tensorD]
- Scikit-tensor for Python [M. Nickel, github.com/mnick/scikit-tensor]
- rTensor for R [J. Li et al., cran.r-project.org/web/packages/rTensor]

3.5. Other methods for tensorial analysis

Sections 3.2 and 3.3 discussed the basic generalizations of PCA to tensor decomposition. Besides CP and Tucker decompositions, other tensorial PCA methods are available including (Lu et al., 2011)

- ➢ Tensor-to-Tensor Projection
  - Two-Dimensional PCA (2DPCA) (Yang et al., 2004)
  - Generalized low rank approximation of matrices (GLRAM) (Ye, 2005)
  - Multilinear PCA (MPCA) (Lu et al., 2008a)
  - Non-negative MPCA (NMPCA) (Panagakis et al., 2010)
  - Bayesian Tensor Analysis (BTA) (Tao et al., 2008b)
  - Incremental Tensor Analysis (ITA) (Sun et al., 2008)
  - Dynamic Tensor Analysis (DTA) (Sun et al., 2006b)
  - Streaming Tensor Analysis (STA) (Papadimitriou et al., 2005)
  - Window-based Tensor Analysis (WTA) (Sun et al., 2006a)
- ➢ Tensor-to-Vector Projection
  - Tensor Rank-One Decomposition (TROD) (Shashua and Levin, 2001)
  - Uncorrelated MPCA (UMPCA) (Lu et al., 2009)

These multilinear PCA methods can be viewed as different forms of tensorial generalization of PCA based on different formulations of the objective functions, constraints, and projection methods. For example, MPCA can be viewed as a special case of the Tucker decomposition, which decomposes the original tensor $\mathcal{A} \in R^{m_1 \times \cdots \times m_d}$ on $d-1$ modes and leaves one mode uncompressed (which typically refers to the sample number). Another example is GLRAM, which only works for a series of matrices. GLRAM can be viewed as a special case of MPCA for three-way arrays.

Besides the unsupervised multilinear subspace learning methods, there are multiple methods for supervised multilinear subspace learning. Tensorial LDA methods for supervised classification include (Lu et al., 2011)

- ➢ Tensor-to-Tensor Projection
  - 2D LDA (2DLDA) (Ye et al., 2005)
  - Discriminant Analysis with Tensor Representation (DATER) (Yan et al., 2005)
  - General Tensor Discriminant Analysis (Tao et al., 2007)
- ➢ Tensor-to-Vector Projection
  - Tensor Rank-One Discriminant Analysis (Tao et al., 2008a)
  - Uncorrelated Multilinear Discriminant Analysis (UMLDA) (Lu et al., 2008b)

For predictive modeling, there are multiple extensions of the commonly used PCR, PLS, and canonical correlation analysis (CCA) (Hardoon et al., 2004) methods, including multilinear PCR (Su et al., 2012), N-way PLS (Bro, 1996), higher order PLS (Zhao et al., 2013), and tensor CCA (Kim et al., 2007; Luo et al., 2015b).

There are also different tensorial extensions to other popular machine learning techniques, such as support tensor machine (Guo et al., 2014; Hao et al., 2013; Xiang et al., 2018) as a generalization of support vector machine. In addition, various types of neural networks can handle tensorial data directly, including convolutional neural networks (Gu et al., 2018), recurrent neural networks (Rumelhart et al., 1986), and tensor net (Oseledets, 2011).

3.6. Applications

The previous sections discussed various types of tensorial data analytics methods. There is no single versatile method suitable for all types of applications because each method has advantages depending on how well its assumptions align with the particular dataset. Besides, the application subjects of different methods are not the same. Since tensorial data analytics has not been widely applied in the chemical and biological manufacturing industries, which method is best for a specific application and how to best use the methods remains an open problem.

With regard to method selection, two points should be carefully assessed: (1) what kind of problem are you trying to solve, and (2) what are the key properties and constraints of each data analytics method. Then, based on the requirements of that specific application, the appropriate method should be selected or developed to construct the model. Here a simple illustrative example is presented for method selection between the CP and Tucker decompositions for spectral analysis.

First of all, as discussed in Section 3.2, the CP decomposition can be viewed as a summation of rank-1 components with no core tensor, which can also be viewed as a super-diagonal core tensor. The columns in the factor matrix are linearly dependent. Besides, under mild conditions, the decomposition is unique. However, in order to enhance the accuracy, prior knowledge of data may be incorporated into constraints to relax the uniqueness condition, such as orthogonality and non-negativity. The CP decomposition is useful for exploratory data analytics and is capable of revealing the natural source of the data. For the Tucker decomposition, the method allows for variable transformation in each mode with orthonormal mode-wise factor matrices and a dense core tensor. The decomposition is non-unique in general. Unlike the CP decomposition, which is typically used for factorizing data into interpretable components, the Tucker decomposition is often used for data compression or to find the subspaces spanned by the fibers. Adding other constraints to the Tucker decomposition, such as non-negativity and sparsity, may help to find a unique solution in the Tucker representation. Therefore, the CP decomposition is often selected for spectral analysis. For example, in Bro (1997), the CP decomposition is used for fluorescence measurements to assess the composition of the samples. The samples contain different amounts of three types of amino acids in buffered water, and a model with three CP components is developed with each component representing a rank-one contribution of one specific amino acid.

Tensor decomposition have been applied to spectral data (Baum et al., 2013; Gu et al., 2016, 2014; Henrion, 1994; Murphy et al., 2014; Nørgaard, 1995), biological signals (Cong et al., 2015; Mahyari et al., 2017; Williams et al., 2018), batch process data (Guo et al., 2010; Hu and Yuan, 2009; 20,Luo et al., 2013, 2015, 2016; Muñoz et al., 2018), and imaging (Dey et al., 2019; Papastergiou et al., 2018; Schultz et al., 2001). While some chemical engineers have recently been learning about tensorial data analytics, the inspection of the literature indicates that the vast majority have not, as suboptimal methods continue to dominate with no acknowledgement or reference to tensorial methods. Given the sparsity of published applications, there remains little known about which methods to apply to which chemical and biological manufacturing processes, and more detailed application studies need to be clearly documented in the open literature before tensorial data analytics becomes widely accepted and consistently applied in the community.

4. Conclusions and future directions

With the development of sensor technologies and wireless networks, different types of sensor data are available in chemical and biological manufacturing processes. It is crucial to effectively utilize those new information streams to further improve the process efficiency, product quality, and process safety. One feature of the new information streams from the manufacturing processes is the presence of higher order tensors. Instead of scalar or vector measurements, a single measurement could be a second-, third-, or

higher order tensors. However, data analytics methods for handling higher order tensorial manufacturing data have not been fully investigated. The vast majority of industrial applications still use traditional statistical learning methods on the unfolded data, which are suboptimal. This article provides an introduction to the opportunities of tensorial data analytics and how these methods work, with the main emphasis on the largest class of methods, multilinear tensor decomposition, which are generalizations of the PCA, PLS, FDA, and related methods commonly applied to matrix data by chemical engineers. Other extensions to tensorial data analytics and their applications in chemical and biological manufacturing processes are also briefly discussed. The goal of this article is to show that tensorial data analytics, as compared with the traditional two-way learning methods, opens up new possibilities for extracting useful process information and is a promising tool for process understanding and optimization.

To enable wide and consistent application of tensorial data analytics to industrial processes, several challenging issues remain to be addressed in more depth:

- Since tensorial data analytics has not been widely applied in chemical and biological manufacturing industries, systematic and in-depth comparison of various tensorial methods for manufacturing application purposes is needed, to provide guidance on which method should be used in any specific situation and how to use the method effectively. Systematic comparison studies should provide rigorous performance analysis and comparisons, and serve as benchmarks for methods development.
- Data pre-processing procedure for tensorial manufacturing data is crucial for consistent application and achieving high model accuracy. For example, tensor models are sensitive to data scale and it is important to scale the tensor data to ensure the same scale for all columns (Louwerse et al., 1999). Another example is pre-processing for higher order spectra for peak alignment and background drift removal.
- New algorithms are needed to further extend the flexibility and accuracy of tensor models. Efficient iterative algorithms, optimal initialization, and methods for automated estimation of critical hyperparameters (e.g., the number of components in tensor decomposition) are needed as the complexity of models increases. Besides, it is important to tailor the model structure to particular data properties of given applications, such as adding additional sparsity and non-negativity constraints to the model.
- Advanced generalization of process data analytics to tensorial analysis is important for complex systems, such as nonlinear, dynamic, and time-varying processes. The model complexity of those advanced methods will be high. Therefore, it is also important to develop approaches to efficiently assess tensorial data properties to inform whether advanced methods are needed for a given application. Another needed extension is probabilistic tensorial modeling (e.g., Bayesian tensor regression (Guhaniyogi et al., 2017)) to incorporate process uncertainties and intrinsic variability for risk assessments, which are critical in industries such as pharmaceutical manufacturing. Prior knowledge such as complex variable interactions and noise distributions can also be incorporated with probabilistic modeling to improve model accuracy and efficiency (Yılmaz and Cemgil, 2010). Finally, tensorial data analytics should be integrated with process knowledge, such as first-principles models and the process structure.

Chemical and biological process systems are in an interesting place right now, in that advances in sensor technologies for higher order tensorial datasets have created opportunities for the generation of vastly larger datasets at much lower costs. Many of these imaging-based sensors are non-contact, which is also highly desirable in applications to many (perhaps most) chemical and biological processes. Chemical processes often involve highly corrosive environments that damage or foul sensors, and it is desirable in biopharmaceutical and biomedical device manufac-turing to reduce the number of potential contact points that could cause product contamination. Such sensors also enable the measurement of spatial concentration field which characterize spatial heterogeneities in chemical and biological processes, from fast combustion processes to food products to freeze-dried biopharmaceutical and cellular products.

The technometricians and computer scientists have also been developing numerous algorithms and freely available software for tensorial data analytics, so those advances can be leveraged by the chemical engineering community to push forward their applications. The combination of new information streams of higher potential value and lower costs with higher order learning methods has significant potential for better and smarter next-generation manufacturing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported by the U.S. Food and Drug Administration. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the financial sponsor.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.compchemeng.2020.107099.

References

Baum, A., Hansen, P.W., Meyer, A.S., Mikkelsen, J.D., 2013. Simultaneous measurement of two enzyme activities using infrared spectroscopy: a comparative evaluation of PARAFAC, TUCKER and N-PLS modeling. Anal. Chim. Acta 790, 14–23. doi:10.1016/j.aca.2013.06.039.

Borden, M., El-Amin, S.F., Attawia, M., Laurencin, C.T., 2003. Structural and human cellular assessment of a novel microsphere-based tissue engineered scaffold for bone repair. Biomaterials 24, 597–609. doi:10.1016/S0142-9612(02)00374-5.

Bro, R., 1999. Exploratory study of sugar production using fluorescence spectroscopy and multi-way analysis. Chemom. Intell. Lab. Syst. 46, 133–147. doi:10.1016/S0169-7439(98)00181-6.

Bro, R., 1997. PARAFAC. Tutorial and applications. Chemom. Intell. Lab. Syst 38, 149–171. doi:10.1016/S0169-7439(97)00032-4.

Bro, R., 1996. Multiway calibration. Multilinear PLS. J. Chemom. 10, 47–61. doi:10.1002/(SICI)1099-128X(199601)10:1⟨47:AID-CEM400⟩3.0.CO;2-C.

Brouckaert, D., De Meyer, L., Vanbillemont, B., Van Bockstal, P.-J., Lammens, J., Mortier, S., Corver, J., Vervaet, C., Nopens, I., De Beer, T., 2018. Potential of Near-Infrared Chemical Imaging as Process Analytical Technology Tool for Continuous Freeze-Drying. Anal. Chem. 90, 4354–4362. doi:10.1021/acs.analchem.7b03647.

Carroll, J.D., Chang, J.-J., 1970. Analysis of individual differences in multidimensional scaling via an n-way generalization of “Eckart-Young” decomposition. Psychometrika 35, 283–319. doi:10.1007/BF02310791.

Chiang, L., Lu, B., Castillo, I., 2017. Big Data Analytics in Chemical Engineering. Annu. Rev. Chem. Biomol. Eng. 8, 63–85. doi:10.1146/annurev-chembioeng-060816-101555.

Chiang, L.H., Russell, E.L., Braatz, R.D., 2000. Fault diagnosis in chemical processes using Fisher discriminant analysis, discriminant partial least squares, and principal component analysis. Chemom. Intell. Lab. Syst. 50, 243–252. doi:10.1016/S0169-7439(99)00061-1.

Cichocki, A., 2014. Era of big data processing: a new approach via tensor networks and tensor decompositions. arXiv:1403.2048.

Cong, F., Lin, Q.-.H., Kuang, L.-.D., Gong, X.-.F., Astikainen, P., Ristaniemi, T., 2015. Tensor decomposition of EEG signals: a brief review. J. Neurosci. Methods 248, 59–69. doi:10.1016/j.jneumeth.2015.03.018.

De Lathauwer, L., De Moor, B., Vandewalle, J., 2000. A multilinear singular value decomposition. SIAM J. Matrix Anal. Appl. 21, 1253–1278. doi:10.1137/S0895479896305696.

Dey, N., Hong, S., Ach, T., Koutalos, Y., Curcio, C.A., Smith, R.T., Gerig, G., 2019. Tensor decomposition of hyperspectral images to study autofluorescence in age-related macular degeneration. Med. Image Anal. 56, 96–109. doi:10.1016/j.media.2019.05.009.

Emteborg, H., Zeleny, R., Charoud-Got, J., Martos, G., Lüddeke, J., Schellin, H., Teipel, K., 2014. Infrared Thermography for Monitoring of Freeze-Drying Processes: instrumental Developments and Preliminary Results. J. Pharm. Sci. 103, 2088–2097. doi:10.1002/jps.24017.

Ge, Z., Song, Z., Gao, F., 2013. Review of Recent Research on Data-Based Process Monitoring. Ind. Eng. Chem. Res. 52, 3543–3562. doi:10.1021/ie302069q.

Geladi, P., Kowalski, B.R., 1986. Partial least-squares regression: a tutorial. Anal. Chim. Acta 185, 1–17. doi:10.1016/0003-2670(86)80028-9.

Grand View Research, 2019. Hyperspectral Imaging Systems Market Size, Share & Trends Analysis Report by Application (Military, Remote Sensing, Medical Diagnostics, Optical Sorting), by Product (Camera, Accessories), and Segment Forecasts, 2019 - 2025. Technical report GVR-2-68038-788-9. Grand View Research.

Gu, H.-W., Wu, H.-L., Li, S.-S., Yin, X.-L., Hu, Y., Xia, H., Fang, H., Yu, R.-Q., Yang, P.-Y., Lu, H.-J., 2016. Chemometrics-enhanced full scan mode of liquid chromatography–mass spectrometry for the simultaneous determination of six co-eluted sulfonylurea-type oral antidiabetic agents in complex samples. Chemom. Intell. Lab. Syst. 155, 62–72. doi:10.1016/j.chemolab.2016.04.001.

Gu, H.-W., Wu, H.-L., Yin, X.-L., Li, Y., Liu, Y.-J., Xia, H., Zhang, S.-R., Jin, Y.-F., Sun, X.-D., Yu, R.-Q., Yang, P.-Y., Lu, H.-J., 2014. Multi-targeted interference-free determination of ten β-blockers in human urine and plasma samples by alternating trilinear decomposition algorithm-assisted liquid chromatography–mass spectrometry in full scan mode: comparison with multiple reaction monitoring. Anal. Chim. Acta 848, 10–24. doi:10.1016/j.aca.2014.08.052.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., Chen, T., 2018. Recent advances in convolutional neural networks. Pattern Recognit 77, 354–377. doi:10.1016/j.patcog.2017.10.013.

Guhaniyogi, R., Qamar, S., Dunson, D.B., 2017. Bayesian tensor regression. J. Mach. Learn. Res 18, 2733–2763.

Gülbakan, B., Barylyuk, K., Zenobi, R., 2015. Determination of thermodynamic and kinetic properties of biomolecules by mass spectrometry. Current Opinion Biotechnology 31, 65–72. doi:10.1016/j.copbio.2014.08.003.

Guo, J., Li, Y., Wang, G., Zeng, J., 2010. Batch Process Monitoring Based on Multilinear Principal Component Analysis. In: 2010 International Conference on Intelligent System Design and Engineering Application, pp. 413–416. doi:10.1109/ISDEA.2010.274.

Guo, X., Huang, X., Zhang, L., Zhang, L., 2014. Support tensor machine with local pixel neighborhood for hyperspectral image classification. In: 2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), pp. 1–4. doi:10.1109/WHISPERS.2014.8077627.

Hübschmann, H.-J., 2015. Handbook of GC-MS: Fundamentals and Applications, 3rd edition Wiley-VCH, Weinheim, Germany doi:10.1007/s00216-015-9282-1.

Hao, Z., He, L., Chen, B., Yang, X., 2013. A Linear Support Higher-Order Tensor Machine for Classification. IEEE Trans. Image Process. 22, 2911–2920. doi:10.1109/TIP.2013.2253485.

Hardoon, D.R., Szedmak, S., Shawe-Taylor, J., 2004. Canonical correlation analysis: an overview with application to learning methods. Neural Comput 16, 2639–2664. doi:10.1162/0899766042321814.

Harshman, R.A., 1970. Foundations of the PARAFAC procedure: models and conditions for an" explanatory" multimodal factor analysis. UCLA Working Papers in Phonetics 16, 1–84.

Henrion, R., 1994. N-way principal component analysis theory, algorithms and applications. Chemom. Intell. Lab. Syst. 25, 1–23. doi:10.1016/0169-7439(93)E0086-J.

Hu, K., Yuan, J., 2009. Batch process monitoring with tensor factorization. J. Process Control 19, 288–296. doi:10.1016/j.jprocont.2008.03.003.

Hu, K., Yuan, J., 2008. Statistical monitoring of fed-batch process using dynamic multiway neighborhood preserving embedding. Chemom. Intell. Lab. Syst. 90, 195–203. doi:10.1016/j.chemolab.2007.10.002.

Hyvärinen, A., Oja, E., 2000. Independent component analysis: algorithms and applications. Neural Networks 13, 411–430. doi:10.1016/S0893-6080(00)00026-5.

Jeffy, F.J., Gugaliya, J.K., Kariwala, V., 2018. Application of Multi-Way Principal Component Analysis on Batch Data, in: 12th International Conference on Control (CONTROL). pp. 414–419. doi:10.1109/CONTROL.2018.8516806

Jackson, J.E., 1991. A User's Guide to Principal Components. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, USA doi:10.1002/0471725331.

Jiang, M., Zhu, Z., Jimenez, E., Papageorgiou, C.D., Waetzig, J., Hardy, A., Langston, M., Braatz, R.D., 2014. Continuous-Flow Tubular Crystallization in Slugs Spontaneously Induced by Hydrodynamics. Cryst. Growth Des. 14, 851–860. doi:10.1021/cg401715e.

Jiang, M., Severson, K., Love, J.C., Madden, H., Swan, P., Zang, L., Braatz, R.D., 2017. Opportunities and challenges of real-time release testing for biopharmaceutical manufacturing. Biotechnol. Bioeng. 114, 2445–2456. doi:10.1002/bit.26383.

Kiers, H.A.L., 2000. Towards a standardized notation and terminology in multiway analysis. J. Chemom. 14, 105–122. doi:10.1002/1099-128X(200005/06)14:3⟨105:AID-CEM582⟩3.0.CO;2-I.

Kim, S., Kim, J., Hwang, M., Kim, M., Jo, S.J., Je, M., Jang, J.E., Lee, D.H., Hwang, J.Y., 2019. Smartphone-based multispectral imaging and machine-learning based analysis for discrimination between seborrheic dermatitis and psoriasis on the scalp. Biomed. Opt. Express. 10, 879–891. doi:10.1364/BOE.10.000879.

Kim, T., Wong, S., Cipolla, R., 2007. Tensor Canonical Correlation Analysis for Action Classification. In: 2007 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8. doi:10.1109/CVPR.2007.383137.

Koek, M.M., van der Kloet, F.M., Kleemann, R., Kooistra, T., Verheij, E.R., Hankemeier, T., 2011. Semi-automated non-target processing in GC × GC–MS metabolomics analysis: applicability for biomedical studies. Metabolomics 7, 1–14. doi:10.1007/s11306-010-0219-6.

Kolda, T.G., Bader, B.W., 2009. Tensor decompositions and applications. SIAM Rev 51, 455–500. doi:10.1137/07070111X.

Kourti, T., 2003. Multivariate dynamic data modeling for analysis and statistical process control of batch processes, start-ups and grade transitions. J. Chemom. 17, 93–109. doi:10.1002/cem.778.

Kruskal, J.B., 1989. In: Rank, decomposition, and Uniqueness For 3-way and N-way arrays, in Multiway Data Analysis, 1. North-Holland Publishing Co., Amsterdam, Netherlands, pp. 7–18.

Kruskal, J.B., 1977. Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. Linear Algebra Appl 18, 95–138. doi:10.1016/0024-3795(77)90069-6.

Lakshminarayanan, S., Gudi, R.D., Shah, S.L., Nandakumar, K., 1996. Monitoring Batch Processes Using Multivariate Statistical Tools: extensions and Practical Issues. IFAC Proc. Vol 29, 6037–6042. doi:10.1016/S1474-6670(17)58648-6.

Lee, J.-M., Yoo, C., Lee, I.-B., 2004. Fault detection of batch processes using multiway kernel principal component analysis. Comput. Chem. Eng. 28, 1837–1847. doi:10.1016/j.compchemeng.2004.02.036.

Liu, J.J., MacGregor, J.F., Duchesne, C., Bartolacci, G., 2005. Flotation froth monitoring using multiresolutional multivariate image analysis. Miner. Eng 18, 65–76. doi:10.1016/j.mineng.2004.05.010.

Louwerse, D.J., Smilde, A.K., Kiers, H.A.L., 1999. Cross-validation of multiway component models. J. Chemom. 13, 491–510. doi:10.1002/(SICI)1099-128X(199909/10)13:5⟨491::AID-CEM537⟩3.0.CO;2-2.

Lu, H., Plataniotis, K.N., Venetsanopoulos, A.N., 2011. A survey of multilinear subspace learning for tensor data. Pattern Recognit 44, 1540–1551. doi:10.1016/j.patcog.2011.01.004.

Lu, H., Plataniotis, K.N., Venetsanopoulos, A.N., 2009. Uncorrelated Multilinear Principal Component Analysis for Unsupervised Multilinear Subspace Learning. IEEE Trans. Neural Networks 20, 1820–1836. doi:10.1109/TNN.2009.2031144.

Lu, H., Plataniotis, K.N., Venetsanopoulos, A.N., 2008a. MPCA: multilinear Principal Component Analysis of Tensor Objects. IEEE Trans. Neural Networks 19, 18–39. doi:10.1109/TNN.2007.901277.

Lu, H., Plataniotis, K.N., Venetsanopoulos, A.N., 2008b. Uncorrelated multilinear discriminant analysis with regularization and aggregation for tensor object recognition. IEEE Trans. Neural Networks 20, 103–123. doi:10.1109/TNN.2008.2004625.

Luo, L., Bao, S., Gao, Z., 2015a. Quality prediction based on HOPLS-CP for batch processes. Chemom. Intell. Lab. Syst. 143, 28–39. doi:10.1016/j.chemolab.2015.02.010.

Luo, L., Bao, S., Gao, Z., Yuan, J., 2013. Batch Process Monitoring with Tensor Global–Local Structure Analysis. Ind. Eng. Chem. Res. 52, 18031–18042. doi:10.1021/ie402355f.

Luo, L., Bao, S., Mao, J., Tang, D., 2016. Quality prediction and quality-relevant monitoring with multilinear PLS for batch processes. Chemom. Intell. Lab. Syst. 150, 9–22. doi:10.1016/j.chemolab.2015.11.004.

Luo, Y., Tao, D., Ramamohanarao, K., Xu, C., Wen, Y., 2015b. Tensor canonical correlation analysis for multi-view dimension reduction. IEEE Trans. Knowl. Data Eng. 27, 3111–3124. doi:10.1109/TKDE.2015.2445757.

MacGregor, J.F., 2004. Data-based latent variable methods for process analysis, monitoring and control. Comput. Aided Chem. Eng. 18, 87–98. doi:10.1016/S1570-7946(04)80085-3.

MacGregor, J.F., Kourti, T., 1995. Statistical process control of multivariate processes. Control Eng. Pract. 3, 403–414. doi:10.1016/0967-0661(95)00014-L.

Mahyari, A.G., Zoltowski, D.M., Bernat, E.M., Aviyente, S., 2017. A Tensor Decomposition-Based Approach for Detecting Dynamic Network States From EEG. IEEE Trans. Biomed. Eng. 64, 225–237. doi:10.1109/TBME.2016.2553960.

Marjanovic, O., Lennox, B., Sandoz, D., Smith, K., Crofts, M., 2006. Real-time monitoring of an industrial batch process. Comput. Chem. Eng. 30, 1476–1481. doi:10.1016/j.compchemeng.2006.05.040.

Mitchell, J.B.O., 2014. Machine learning methods in chemoinformatics. Wiley Interdiscip. Rev. Comput. Mol. Sci. 4, 468–481. doi:10.1002/wcms.1183.

Moe, S., Rustad, A.M., Hanssen, K.G., 2018, Eds. In: Bramer, M., Petridis, M. (Eds.). Springer International Publishing, Cham, pp. 250–265. doi:10.1007/978-3-030-04191-5_23.

Muñoz, C.A., Telen, D., Nimmegeers, P., Impe, J.Van, 2018. Feature extraction for batch process monitoring and fault detection via simultaneous data scaling and training of tensor based models. IFAC-PapersOnLine 51, 433–440. doi:10.1016/j.ifacol.2018.09.613.

Murphy, K.R., Stedmon, C.A., Wenig, P., Bro, R., 2014. OpenFluor–an online spectral library of auto-fluorescence by organic compounds in the environment. Anal. Methods 6, 658–661. doi:10.1039/C3AY41935E.

Næs, T., Martens, H., 1988. Principal component regression in NIR analysis: viewpoints, background details and selection of components. J. Chemom. 2, 155–167. doi:10.1002/cem.1180020207.

Nørgaard, L., 1995. A multivariate chemometric approach to fluorescence spectroscopy. Talanta 42, 1305–1324. doi:10.1016/0039-9140(95)01586-Z.

Oseledets, I.V., 2011. Tensor-train decomposition. SIAM J. Sci. Comput. 33, 2295–2317. doi:10.1137/090752286.

Ou-Yang, C.-F., Chen, Y.-J., Lee, C.-T., Chi, K.-H., Lin, N.-H., Wang, J.L., 2018. Analyzing Organic Constituents on Atmospheric Particulate Matters by Comprehensive Two-Dimensional Gas Chromatography-Mass Spectrometry (GCxGC-MS), in: AGU Fall Meeting Abstracts.

Panagakis, Y., Kotropoulos, C., Arce, G.R., 2010. Non-Negative Multilinear Principal Component Analysis of Auditory Temporal Modulations for Music Genre Classification. IEEE Trans. Audio. Speech. Lang. Process. 18, 576–588. doi:10.1109/TASL.2009.2036813.

Papadimitriou, S., Sun, J., Faloutsos, C., 2005. Streaming Pattern Discovery in Multiple Time-Series. In: VLDB '05: Proceedings of the 31st International Conference on Very Large Data Bases, pp. 697–708.

Papalexakis, E.E., Faloutsos, C., Sidiropoulos, N.D., 2016. Tensors for data mining and data fusion: models, applications, and scalable algorithms. ACM Trans. Intell. Syst. Technol. 8, 1–44. doi:10.1145/2915921.

Papastergiou, T., Zacharaki, E.I., Megalooikonomou, V., 2018. Tensor Decomposition for Multiple-Instance Classification of High-Order Medical Data. Complexity 2018, 8651930. doi:10.1155/2018/8651930.

Patel, K.N., Patel, J.K., Patel, M.P., Rajput, G.C., Patel, H.A., 2010. Introduction to hyphenated techniques and their applications in pharmacy. Pharm. Methods 1, 2–13. doi:10.4103/2229-4708.72222.

Qin, S.J., 2012. Survey on data-driven industrial process monitoring and diagnosis. Annu. Rev. Control 36, 220–234. doi:10.1016/j.arcontrol.2012.09.004.

Rabanser, S., Shchur, O., Günnemann, S., 2017. Introduction to tensor decompositions and their applications in machine learning. arXiv:1711.10781.

Rapin, J., Souloumiac, A., Bobin, J., Larue, A., Junot, C., Ouethrani, M., Starck, J.-L., 2016. Application of non-negative matrix factorization to LC/MS data. Signal Processing 123, 75–83. doi:10.1016/j.sigpro.2015.12.014.

Rogers, R.S., Nightlinger, N.S., Livingston, B., Campbell, P., Bailey, R., Balland, A., 2015. Development of a quantitative mass spectrometry multi-attribute method for characterization, quality control testing and disposition of biologics. MAbs 7, 881–890. doi:10.1080/19420862.2015.1069454.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature 323, 533–536. doi:10.1038/323533a0.

Salazar-Vazquez, J., Mendez-Vazquez, A., 2020. A plug-and-play Hyperspectral Imaging Sensor using low-cost equipment. HardwareX 7, e00087.

Schultz, R.A., Nielsen, T., Zavaleta, J.R., Ruch, R., Wyatt, R., Garner, H.R., 2001. Hyperspectral imaging: a novel approach for microscopic analysis. Cytometry 43, 239–247. doi:10.1002/1097-0320(20010401)43:4⟨239::AID-CYTO1056⟩3.0.CO;2-Z.

Seger, C., Godejohann, M., Tseng, L.-H., Spraul, M., Girtler, A., Sturm, S., Stuppner, H., 2005. LC-DAD-MS/SPE-NMR Hyphenation. A Tool for the Analysis of Pharmaceutically Used Plant Extracts: Identification of Isobaric Iridoid Glycoside Regioisomers from Harpagophytum procumbens. Anal. Chem. 77, 878–885. doi:10.1021/ac048772r.

Severson, K., Chaiwatanodom, P., Braatz, R.D., 2016. Perspectives on Process Monitoring of Industrial Systems. Annu. Rev. Control 42, 190–200. doi:10.1016/j.ifacol.2015.09.646.

Shashua, A., Levin, A., 2001. Linear image coding for regression and classification using the tensor-rank principle, in: proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001. doi:10.1109/CVPR.2001.990454, I–I.

Shi, X., Wei, X., Koo, I., Schmidt, R.H., Yin, X., Kim, S.H., Vaughn, A., McClain, C.J., Arteel, G.E., Zhang, X., Watson, W.H., 2014. Metabolomic Analysis of the Effects of Chronic Arsenic Exposure in a Mouse Model of Diet-Induced Fatty Liver Disease. J. Proteome Res. 13, 547–554. doi:10.1021/pr400719u.

Sidiropoulos, N.D., Lathauwer, L.De, Fu, X., Huang, K., Papalexakis, E.E., Faloutsos, C., 2017. Tensor Decomposition for Signal Processing and Machine Learning. IEEE Trans. Signal Process. 65, 3551–3582. doi:10.1109/TSP.2017.2690524.

Su, Y., Gao, X., Li, X., Tao, D., 2012. Multivariate Multilinear Regression. IEEE Trans. Syst. Man. Cybern. Part B 42, 1560–1573. doi:10.1109/TSMCB.2012.2195171.

Sun, J., Papadimitriou, S., Philip, S.Y., 2006a. Window-based tensor analysis on high-dimensional and multi-aspect streams. In: Sixth International Conference on Data Mining (ICDM'06). IEEE, pp. 1076–1080. doi:10.1109/ICDM.2006.169.

Sun, J., Tao, D., Faloutsos, C., 2006b. Beyond streams and graphs: dynamic tensor analysis. In: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 374–383. doi:10.1145/1150402.1150445.

Sun, J., Tao, D., Papadimitriou, S., Yu, P.S., Faloutsos, C., 2008. Incremental tensor analysis: theory and applications. ACM Trans. Knowl. Discov. from Data 2, 1–37. doi:10.1145/1409620.1409621.

Tao, D., Li, X., Wu, X., Maybank, S., 2008a. Tensor Rank One Discriminant Analysis—A convergent method for discriminative multilinear subspace selection. Neurocomputing 71, 1866–1882. doi:10.1016/j.neucom.2007.08.036.

Tao, D., Li, X., Wu, X., Maybank, S.J., 2007. General Tensor Discriminant Analysis and Gabor Features for Gait Recognition. IEEE Trans. Pattern Anal. Mach. Intell. 29, 1700–1715. doi:10.1109/TPAMI.2007.1096.

Tao, D., Sun, J., Shen, J., Wu, X., Li, X., Maybank, S.J., Faloutsos, C., 2008b. Bayesian tensor analysis. In: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence). IEEE, pp. 1402–1409. doi:10.1109/IJCNN.2008.4633981.

Tucker, L.R., 1966. Some mathematical notes on three-mode factor analysis. Psychometrika 31, 279–311. doi:10.1007/BF02289464.

Tucker, L.R., 1963. Implications of factor analysis of three-way matrices for measurement of change. Probl. Meas. Chang. 15, 122–137.

Wang, R.C., Edgar, T.F., Baldea, M., Nixon, M., Wojsznis, W., Dunia, R., 2015. Process fault detection using time-explicit Kiviat diagrams. AIChE J 61, 4277–4293. doi:10.1002/aic.15054.

Welling, M., 2009. Fisher Linear Discriminant Analysis. Technical Report, Department of Computer Science. University of Toronto, Canada.

Williams, A.H., Kim, T.H., Wang, F., Vyas, S., Ryu, S.I., Shenoy, K.V., Schnitzer, M., Kolda, T.G., Ganguli, S., 2018. Unsupervised Discovery of Demixed, Low-Dimensional Neural Dynamics across Multiple Timescales through Tensor Component Analysis. Neuron 98, 1099–1115. doi:10.1016/j.neuron.2018.05.015.

Winnike, J.H., Wei, X., Knagge, K.J., Colman, S.D., Gregory, S.G., Zhang, X., 2015. Comparison of GC-MS and GC×GC-MS in the Analysis of Human Serum Samples for Biomarker Discovery. J. Proteome Res. 14, 1810–1817. doi:10.1021/pr5011923.

Xiang, Y., Jiang, Q., He, J., Jin, X., Wu, L., Yao, S., 2018. The Advance of Support Tensor Machine. In: 2018 IEEE 16th International Conference on Software Engineering Research. Management and Applications (SERA), pp. 121–128. doi:10.1109/SERA.2018.8477228.

Yan, S., Xu, D., Yang, Q., Zhang, L., Tang, X., Zhang, H.-J., 2005. Discriminant analysis with tensor representation. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 1, pp. 526–532. doi:10.1109/CVPR.2005.131.

Yang, J., Zhang, D., Frangi, A.F., Yang, J.-Y., 2004. Two-dimensional PCA: a new approach to appearance-based face representation and recognition. IEEE Trans. Pattern Anal. Mach. Intell. 26, 131–137. doi:10.1109/TPAMI.2004.1261097.

Ye, J., 2005. Generalized Low Rank Approximations of Matrices. Mach. Learn. 61, 167–191. doi:10.1007/s10994-005-3561-6.

Ye, J., Janardan, R., Li, Q., 2005. Two-dimensional linear discriminant analysis, in: adv. Neural. Inf. Process. Syst 17, 1569–1576.

Yılmaz, Y.K., Cemgil, A.T., 2010. Probabilistic Latent Tensor factorization, in: Latent Variable Analysis and Signal Separation. Springer-Verlag, Berlin-Heidelberg, pp. 346–353.

Yu, J., Qin, S.J., 2009. Multiway Gaussian Mixture Model Based Multiphase Batch Process Monitoring. Ind. Eng. Chem. Res. 48, 8585–8594. doi:10.1021/ie900479g.

Zhao, Q., Caiafa, C.F., Mandic, D.P., Chao, Z.C., Nagasaka, Y., Fujii, N., Zhang, L., Cichocki, A., 2013. Higher Order Partial Least Squares (HOPLS): a Generalized Multilinear Regression Method. IEEE Trans. Pattern Anal. Mach. Intell. 35, 1660–1673. doi:10.1109/TPAMI.2012.254.