The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM)
ISSN: 2602-3199

The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), 2023

Volume 23, Pages 93-99

ICRETS 2023: International Conference on Research in Engineering, Technology and Science

On the Development of the Fluorescence Excitation-Emission Etalon Matrix Algorithm of Wine

Miranda Khajishvili
Batumi Shota Rustaveli State University

Jaba Shainidze
Batumi Shota Rustaveli State University

Kakha Makharadze
Batumi Shota Rustaveli State University

Nugzar Gomidze
Batumi Shota Rustaveli State University

Abstract: Our research provides for the analysis of different types of Georgian wine based on 3D fluorescence spectroscopy (3DF) using the Black Comet (200-950 nm) spectrometer manufactured by StellarNet. In this method, the 3D fluorescence signal is divided into a fixed number of statistical components. For each type of wine, a 3D database is strictly defined, which we conventionally call references. The etalon describe the excitation/emission spectra in detail. The advantage of the 3DF method compared to other statistical methods, such as peak component analysis (PCA), lies in the uniqueness of the unfolding of the spectra. The fluorescence spectra of the wine will be further analyzed by peak component analysis (PCA). After performing the PCA analysis, in order to reduce the number of tolerant etalon, we used the tolerant etalon sample (TES) comparison analysis, thus determining how tolerant the researched wine sample is to this or that specific etalon.

Keywords: 3D fluorescence spectroscopy, Peak component analysis, Wine analysis, Georgian wine, Tolerant etalon sample

Introduction

The combination of 3D fluorescence spectroscopy (3DF) and peak component analysis (PCA) has been used in various fields, including chemistry, biology, environmental science, and food analysis.The combination of 3D fluorescence spectroscopy (3DF) and peak component analysis (PCA) presents a powerful tool for quality control and authentication in the wine industry, particularly for Georgian wine. Here are some key points to highlight about how this combination can be beneficial:

- 3DF is known for its high sensitivity, allowing it to detect even subtle differences in fluorescence patterns. By employing PCA, which helps in identifying specific spectral features, the method becomes even more discriminative. This means that even small variations in the fluorescence spectra of different wine samples can be distinguished and analyzed effectively.
- The unique fluorescence patterns obtained through 3DF can reveal a wealth of information about the chemical composition and structural properties of the wine. PCA further enhances this characterization

by extracting the most significant spectral components. As a result, a comprehensive and detailed picture of each wine type's fluorescence profile is obtained.

- The use of well-defined references (etalons) for each type of Georgian wine enables the establishment of a reliable database. This database acts as a benchmark against which new wine samples can be compared. Any deviation from the known fluorescence profiles can raise a flag for further investigation, helping to identify potential adulteration or counterfeit products.
- With a robust database of fluorescence spectra for various Georgian wines, it becomes easier to trace the origin of a specific wine sample. This is especially valuable for safeguarding wines with geographical indications, as it helps verify whether a wine truly originates from the claimed region.
- One significant advantage of 3DF and PCA is that they are non-destructive techniques. This means that the wine samples do not undergo any chemical alteration during analysis, making it possible to preserve the integrity of the samples for further studies or sensory evaluations.
- Once the reference database is established, the analysis of new wine samples becomes more efficient in terms of time and cost. The comparison with etalons can quickly provide information about the wine's authenticity and potential quality.
- The combination of 3DF and PCA not only benefits wine producers and regulators but also contributes to scientific research. It offers insights into the variability of wine compositions and how different factors, such as grape variety, terroir, and winemaking techniques, influence the fluorescence patterns.

Literature Review

The algorithm for processing the fluorescence excitation-emission matrix for the classification of Argentine white wine is presented in (Azcarate et al., 2015). The effectiveness of using the TES method in wine classification lies in the fact (Wold, 1976) that the types of molecules (such as polyphenols, vitamins, amino acids) and the amount depend on the specific type and maturity of the wine, as well as the wine technology (Urbano et al., 2006; Airado-Rodriguez et al., 2011).

The study includes fluorescence spectroscopy excitation/emission matrix (AEM) analysis, peak component analysis (PCA) and tolerance etalon sample (TES) comparison analysis method development and modeling according to wine product variety and origin. About 100 samples of four types of white Georgian wine were taken. The methodology chosen by us is based on the one hand on the hardware complex, which was gradually modernized by our group (Gomidze et al., 2012; Gomidze et al., 2014; Gomidze et al., 2016; Gomidze et al., 2018), on the other hand on the development of new analytical approaches (Khajisvili et al., 2021) that are quite acceptable to be used in typical laboratory control of food products and beverages. For analyses 3D spectra it is known techniques that are specifically designed for spectroscopic data analysis, such as Multivariate Curve Resolution (MCR), Parallel Factor Analysis (PARAFAC), or Multivariate Analysis of Variance (MANOVA).

Description of the Experiment

Fluorescence spectra were recorded using a Black Comet (200-950 nm) spectrometer manufactured by StellarNet. LED lamps of different frequencies were used as light sources. A wine sample of 100 µl is placed in a quartz cuvette and the spectra are recorded at room temperature. The number of scans is determined from the same experimental measurement to exclude drift effects on the sample. At the beginning of each experiment, the standard is calibrated. The excitation wavelength range is between 250-500 nm, and the emission wavelength is between 275-600 nm. Measurements are performed at different excitation wavelengths with a 5 nm bias. The total time to scan a sample is approximately 10 minutes. Measurements were performed over a short period of time (10-15 days), thereby minimizing the influence of atmospheric effects and instrumental fluctuations (eg. lamp intensity fluctuations). PCA was performed for descriptive analysis of spectral features and TES modeling of analog classes will be used to classify these data. SpectraWiz and LAbView software were used for graphical visualization of the spectra. Data recording and processing were performed in MS Excel and MySQL.

Description of the Theory and Method

Given a dataset with n observations (data points) and 'm' features (variables), we first preprocess the data by centering the variables. Let's assume the centered dataset as x_C. The covariance matrix (C) of the centered data

x_C is a square $m \times m$ matrix where each element $C(i,j)$ represents the covariance between the i-th and j-th variables. The formula for the covariance between two variables x_i and x_j is given as:

$$C(x_i, x_j) = \sum_{n=1}^{N} \frac{(x_i - \bar{x_i})(x_j - \bar{x_j})}{n-1}$$

Next, we perform eigendecomposition on the covariance matrix C to find its eigenvalues (λ) and corresponding eigenvectors (v). The eigendecomposition equation is:

$$C\ v = \lambda\ v$$

Where v is the eigenvector, and λ is the corresponding eigenvalue. The eigenvectors are sorted in descending order based on their corresponding eigenvalues (in decreasing order of importance). This sorting will help us select the most significant principal components. After sorting the eigenvectors, we select the k eigenvectors with the highest corresponding eigenvalues to form the principal components. k is the number of dimensions we want to reduce the data to, and it is typically chosen based on a desired level of explained variance or the number of significant components required for analysis.

We create a matrix W by stacking the k selected eigenvectors as columns. The matrix W will have dimensions $m \times k$. We then transform the original centered data x_C into a reduced-dimensional space by multiplying it with W. This transformation gives us the reduced dataset x_{PCA}:

$$x_{PCA} = x_C \times \mathrm{W}$$

The resulting dataset x_{PCA} contains the principal components, which are the linear combinations of the original features. Each principal component captures a different direction of maximum variance in the data. The first principal component has the highest variance, and subsequent components capture progressively less variance.
The principal components can be used for data visualization, feature selection, or as input for other machine learning algorithms. They represent new axes in the reduced-dimensional space that provide a more concise representation of the original data while preserving most of the variance.

PARAFAC (Parallel Factor Analysis), also known as CANDECOMP or PARAFAC (Canonical Decomposition), is a multivariate statistical method used for analyzing multi-way arrays or tensors. It is a powerful technique for decomposing higher-order data structures into a set of component matrices and capturing the underlying latent factors that explain the observed data. PARAFAC deals with multi-way data arrays (tensors) rather than simple matrices. A tensor is a generalization of a matrix and can be thought of as an n-dimensional array. For example, a matrix is a 2-way tensor (rows and columns), and a 3-way tensor has three modes. PARAFAC aims to approximate the original tensor X with a lower-rank approximation, represented by three component matrices (A, B, and C). The rank of the PARAFAC model is the number of components used to approximate the original tensor.

For a 3-way tensor X with dimensions $I\ x\ J\ x\ K$, the PARAFAC model can be represented as:

$$X = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} A_i B_j C_k$$

Here, A is a matrix of dimension I x rank, B is a matrix of dimension J x rank, and C is a matrix of dimension K x rank.

The goal is to find the factor matrices A, B, and C such that their product approximates the original tensor X as closely as possible. This can be achieved through methods like alternating least squares (ALS) or non-linear optimization techniques. Once the factor matrices A, B, and C are obtained, they provide insights into the underlying latent factors that explain the structure of the original tensor. Each row of A, B, and C represents a specific "mode" or factor of the data. These factors are often interpreted based on the context of the problem.
One common problem with PARAFAC is determining the appropriate rank (number of components) for the model. Choosing the right rank is essential because too low a rank may lead to an oversimplified representation of the data, while too high a rank may lead to overfitting and capturing noise.

Problem Statement: Given a multi-way tensor dataset, we want to apply the PARAFAC method to extract underlying latent factors. However, we are unsure about the appropriate rank for the model. How can we determine the best rank that adequately represents the data without overfitting or underfitting?

In order to model excitation-emission data, the excitation/emission wavelengths of N samples must be placed in a three-dimensional array of size $i \times j \times k$, where i is the number of samples, j is the number of emission wavelengths, k is the number of excitation wavelengths:

$$x_{ijk} = \sum_{n=1}^{N} a_{in} b_{jn} c_{kn} + e_{ijk}$$

where N is the number of samples. Matrix A with elements a_{in} in is conventionally called stroboscopic, and matrices B and C with elements b_{jn} and c_{kn} respectively are called emission and excitation loads. The e_{ijk} elements represent the deviation from the statistical mean for each sample. x_{ijk} - practically represents the sum of signals received from fluorophores.

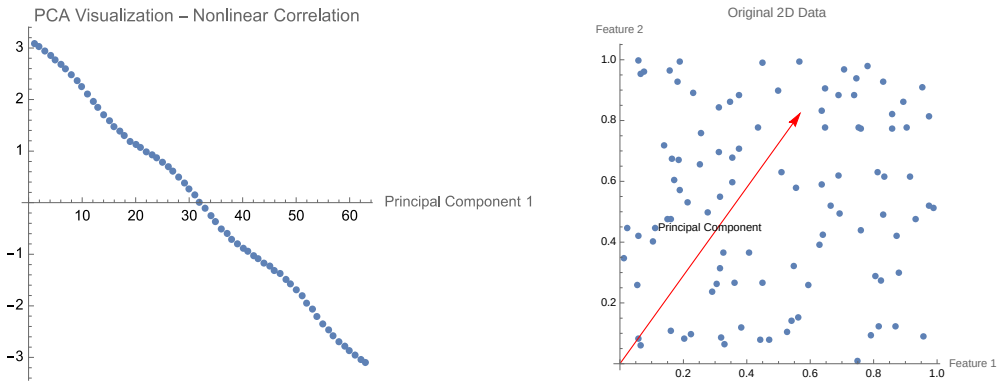
Within this model, the elements of the matrix a_{in} can be interpreted as the concentration of fluorophores n in sample i. The load matrix elements b_{jn} are the basis for the scaled spectrum estimation of the n-th fluorophore at the j-frequency, while the c_{kn} matrix element is proportional to the absorption coefficient of the fluorophore at the k-th frequency (Andersen C. M. & Bro, R., 2003). Excitation/emission array (AEM) coefficients cannot be negative.

Standardization is done by grouping the data for each variety of grapes and their geographical origin, for example West Georgia or East Georgia. In order to perform the analysis of the fluorescence signals of the main components at a fixed - specific k-frequency of excitation, it is necessary to form a two-dimensional matrix $i \times j$ from the main initial array x_{ij}. The goal of PCA analysis is to reduce the rank of the matrix by eliminating redundant members from the data array. For this, we need to find an array of new elements of the matrix in the j -dimensional frequency domain (space) and project the data onto it. The axes of the matrix should be selected so that the data

$$x_{ij} = \sum_{r=1}^{R} t_{ir} p_{jr} + e_{ij}$$

have maximum variance. It turns out that the unit vectors of this new array are precisely the eigenvectors of the xx^T matrix.

Results



a) b)

Figure 1. a) 2D data with non-linear correlation and the execution of PCA to reduce the dimensionality to 1D, b) 2D plot that visualizes the original 2D data and the first principal component vector obtained from PCA

With PCA analysis, we will build tables and graphs for the sample of a specific group. Figure 1a shows the generation of 2D data with non-linear correlation and the execution of PCA to reduce the dimensionality to 1D. We visualize data points in a reduced one-dimensional space. Figure 1b shows a 2D plot that visualizes the original 2D data and the first principal component vector obtained from PCA. In Figure 1b visualized the original 2D data as scattered points and plot the first principal component vector obtained from PCA as a red arrow. The arrow represents the direction of maximum variance in the data, which corresponds to the first principal component. The plot also includes a label indicating that it represents the principal component.

In Figure 2 given 3D plot to visualize the original 3D data, the principal component vectors obtained from PCA, and the data points projected onto the principal component subspace. In this 3D plot, we visualize the original 3D data as scattered points, plot the first two principal component vectors obtained from PCA. The data points projected onto the principal component subspace as green points.

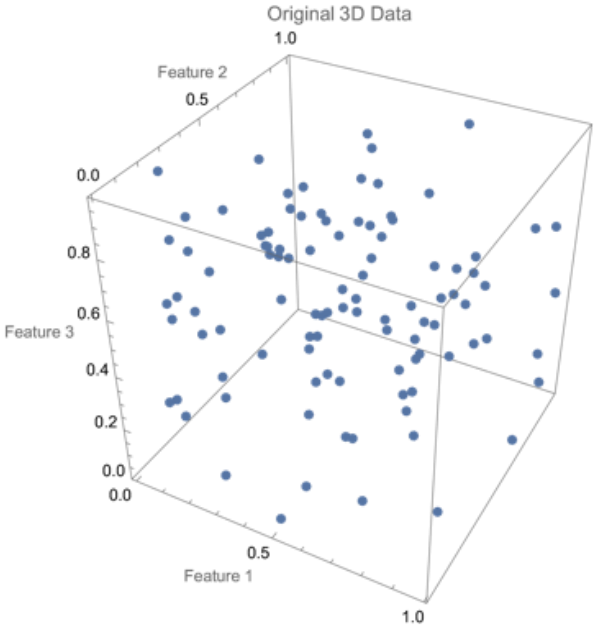


Figure 2. 3D plot to visualize the original 3D data, the principal component vectors obtained from PCA

Figure 3 shows emission spectra of Georgian white wine for Tsolikauri (blue) and Rkatsiteli (red). Three-dimensional graphs account for excitation/emission wavelengths.

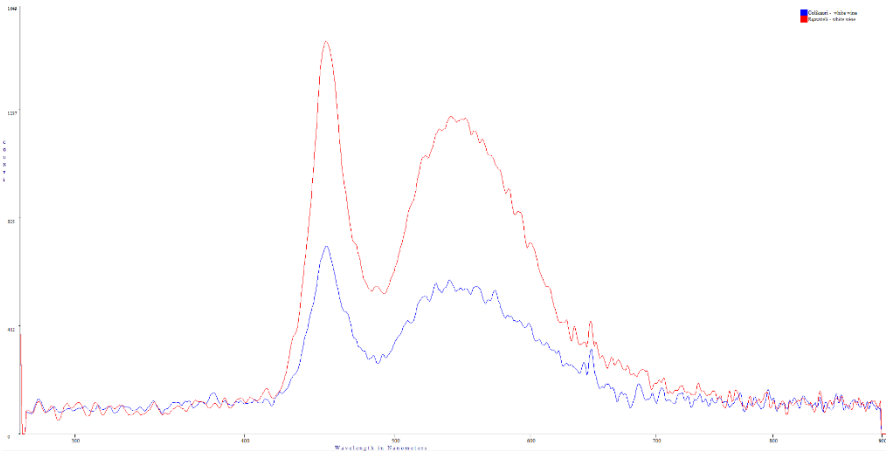


Figure 3. Emission spectra of Georgian white wine: Tsolikauri (blue), Rkatsiteli (red)

Thus, in order to classify wine by type, it is necessary to select a database of a subgroup, which can be called a study group, which in the case of a known type of wine includes many (several dozen) samples. PCA analysis is performed independently on a subset of each known species. This data gives us a spatial picture with the main components. Obviously, the surface of this spatial image is very sensitive to the data of the selected subgroup, so the next step is to analyze and exclude the components with high dispersion from the statistical average.

which is carried out by comparing the selected group and the main components. More precisely by determining the distance between the peaks. If the distance is smaller than some critical value s_0, then the research sample belongs to the class used for comparison. Once such a procedure is completed, a model will be created for each variety of wine, which can be used as a etalons.

Conclusion

The combination of 3DF and PCA allows for a detailed assessment of the unique fluorescence patterns present in different types of Georgian wine. This can be used as a reliable method for quality control and authentication, helping to identify any adulteration or counterfeit products in the market. By comparing the fluorescence spectra of a given wine sample with the well-defined references (etalons), we can verify its authenticity and origin.The fluorescence spectra of wine can be influenced by factors such as grape variety, soil composition, and climate conditions. The 3DF method, combined with PCA and etalon references, can potentially be used to establish a link between the fluorescence patterns and the geographical origin of the wine. This could be valuable for wine producers aiming to protect and promote wines with specific geographical indications.Different vintages of the same wine type can exhibit variations in their fluorescence properties due to varying environmental conditions and winemaking processes. The 3DF analysis, along with PCA and etalon references, might help discern these subtle differences and aid in distinguishing between wines from different years.Some wines are known for their unique characteristics and premium quality, which are often reflected in their distinct fluorescence profiles. Utilizing the 3DF method with PCA analysis can help classify wines into different quality grades based on their fluorescence patterns. This information can be valuable for consumers, sommeliers, and wine enthusiasts when making purchasing decisions.Understanding the fluorescence properties of different wines can provide insights into the chemical composition and structural changes during the winemaking process. This knowledge may assist winemakers in optimizing their production techniques and ensuring consistent quality in the final product.

The 3DF technique can provide a wealth of data on the complex chemical composition of wines. Researchers can use this information to study the presence of various compounds and their interactions, contributing to a deeper understanding of wine chemistry and its influence on wine characteristics. Overall, the combination of 3DF, PCA, and etalon references in the analysis of Georgian wine offers a powerful and sophisticated approach to gain valuable insights into the unique fluorescence properties of different wine types. As this technology evolves and becomes more established, it has the potential to revolutionize the field of wine analysis and enhance various aspects of the wine industry.

Scientific Ethics Declaration

The authors declare that the scientific ethical and legal responsibility of this article published in EPSTEM journal belongs to the authors.

Acknowledgements or Notes

* This article was presented as an oral presentation at the International Conference on Research in Engineering, Technology and Science (www.icrets.net) held in Budapest/Hungary on July 06-09, 2023.

* This work was supported by the 2023 Competition for Targeted Scientific Research Projects titled “On the Development of the Fluorescence Excitation-Emission Etalon Matrix Algorithm of Wine” by Batumi Shota Rustaveli State University. The project manager of this project is Dr. Miranda Khajishvili.

References

Airado–Rodríquez D., Durán-Merás, I., Galeano-Díaz, T. & Wold, J. P. (2011). Front-face fluorescence spectroscopy: A new tool for control in the wine industry. *J. Food Compos. Anal, 24*(2), 257–264.

Andersen C. M. & Bro, R. (2003). Practical aspects of PARAFAC modeling of fluorescence excitation – emission data. *J. Chemometer, 17*, 200–215.

Azcarate S. M. et al. (2015). Modeling excitation–emission fluorescence matrices with pattern recognition algorithms for classification of Argentine white wines according grape variety. *Food Chem*, 184, 214–219.

Gomidze N., Jabnidze I., Makharadze K., Khajishvili M., Shashikadze Z., Surmanidze Z., Surmanidze I. (2012). Numerical analyses of fluorescence characteristics of watery media via laser spectroscopy *Method. Journal of Advanced Materials Research, 590,* 206-211.

Gomidze N.K., Makharadze K.A., Khajishvili M.R., Jabnidze I.N., & Shashikadze Z.K. (2014). Some issues of fluorescence characteristics aqueous media via diagnosis of laser spectroscopy method. *International Journal of Engineering, Science and Innovative Technology, 3*(3), 142-152.

Gomidze N.K., Jabnidze I.N., & Surmanidze Z.J. (2016). Stroboscopic method of fluorescence analyses of optically solid media. *IEEE 7th International Conference on Advanced Optoelectronics and Lasers (CAOL)*. September 12-15, Odessa, Ukraine, (pp. 34-36).

Gomidze N., Shainidze J., Shengelia G., & Turmanidze R. (2018). To the problems of fluorescence excitation spectrums. *Machines. Technologies. Materials., 12*(7), 279-282.

Khajisvili M.R., Gomidze N.K, & Shainidze J.J. (2021). 3D fluorescence spectroscopy of liquid media via internal reference method. In *International Conference on Global Research and Education (pp. 59-71).* Singapore: Springer Singapore.

Urbano M., Luque de Castro, M. D., Pérez, P. M., García – Olmo, J. & Gómez – Nieyo, M. A. (2006). Ultraviolet – visible spectroscopy and pattern recognition methods for differentiation and classification of wines. *Food Chem, 97*(1), 166–175.

Wold S. (1976). Pattern recognition by means of disjoint principal components models. *Pattern Recognit, 8,* 127–139.

Author Information

Miranda Khajishvili
Batumi Shota Rustaveli State University
35/32, Ninoshvili/Rustaveli str., Batumi,
Georgia Contact e-mail: *mirandukht@gmail.com*

Jaba Shainidze
Batumi Shota Rustaveli State University
35/32, Ninoshvili/Rustaveli str., Batumi, Georgia

Kakha Makharadze
Batumi Shota Rustaveli State University
35/32, Ninoshvili/Rustaveli str., Batumi, Georgia

Nugzar Gomidze
Batumi Shota Rustaveli State University
35/32, Ninoshvili/Rustaveli str., Batumi, Georgia

To cite this article:

Khajishvili, M., Shainidze, J., Makharadze, K. & Gomidze, N. (2023). On the development of the fluorescence excitation-emission etalon matrix algorithm of wine. *The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), 23,* 93-99.