# Multivariate analysis on fused hyperspectral datasets within Cultural Heritage field

Di Benedetto, Alessia; de Almieda Nieto, Luìs Manuel; Candeo, Alessia; Valentini, Gianluca; Comelli, Daniela; Alfeld, Matthias

**DOI**
10.1051/epjconf/202430914007

**Publication date**
2024

**Document Version**
Final published version

**Published in**
EPJ Web of Conferences

**Citation (APA)**
Di Benedetto, A., de Almieda Nieto, L. M., Candeo, A., Valentini, G., Comelli, D., & Alfeld, M. (2024). Multivariate analysis on fused hyperspectral datasets within Cultural Heritage field. *EPJ Web of Conferences*, *309*, Article 14007. https://doi.org/10.1051/epjconf/202430914007

# Multivariate analysis on fused hyperspectral datasets within Cultural Heritage field

*Alessia* Di Benedetto[1,2*], *Luìs Manuel* de Almieda Nieto[2], *Alessia* Candeo[1], *Gianluca* Valentini[1], *Daniela* Comelli[1] and *Matthias* Alfeld[2]

[1]Politecnico di Milano, Physics Department, Piazza Leonardo da Vinci 32, 20133 Milano
[2]Delft University of Technology, Department of Material Science and Engineering, Mekelweg 2, 2628 CD Delft

**Abstract.** This work introduces a novel method to multivariate analysis applied to fused hyperspectral datasets in the field of Cultural Heritage (CH). Hyperspectral Imaging is a well-established approach for the non-invasive examination of artworks, offering insights into their composition and conservation status. In CH field, a combination of hyperspectral techniques is usually employed to reach a comprehensive understanding of the artwork. To deal with hyperspectral data, multivariate statistical methods are essential due to the complexity of the data. The process involves factorizing the data matrix to highlight components and reduce dimensionality, with techniques such as Non-negative Matrix Factorization (NMF) gaining prominence. To maximize the synergies between multimodal datasets, the fusion of hyperspectral datasets can be coupled with multivariate analysis, with potential applications in CH. In this work, I will show examples of this approach with different combinations of datasets, including reflectance and transmittance spectral imaging, Fluorescence Lifetime Imaging and Time-Gated Hyperspectral Imaging, and Raman and fluorescence spectroscopy micro-mapping.

## 1 Introduction

The scientific investigation of Cultural Heritage (CH) objects has become indispensable for gaining valuable insights on both their composition and conservation status. Among the different approaches, Hyperspectral Imaging (HSI) has emerged as a highly effective tool, providing insights into the study of diverse artifacts and historical objects.
Cultural Heritage artifacts are extremely complex from a material perspective, due to the presence of multiple chemical species, including degradation products, heterogeneously distributed throughout their structure, from surface to inner layers. To assess their conservation status and evaluate the compounds present, a multi-analytical approach is typically employed by collecting a variety of spectroscopy and imaging data of the artwork. Indeed, the analysis encompasses a wide range of in-situ and laboratory techniques that, given the value of the objects under investigations, are preferably non-destructive and non-invasive [1,2].
Ultimately, the adoption a multimodal approach holds great promise in significantly enhancing our ability to characterize the materials present in works of art and achieve a comprehensive information about artworks.

## 2 Multivariate analyses

Hyperspectral datasets consist of a series of images, each representing a different wavelength, and each pixel of a spectral data-cube contains a spectrum that is a combination of multiple spectra originating from different compounds present in the point. Due to the high number of spectra and their copresence in a single point, multivariate statistical methods are essential to reduce dataset dimensionality and retrieve the principal components.

Therefore, multivariate analysis techniques aim to data matrix factorization, in order to eliminate redundancies, highlighting components and, ultimately, obtain a representation of the data with reduced dimensionality. This approach identifies spectral endmembers and deconstruct each spectrum into a linear combination of these components, weighted based on their specific abundances [3].
To achieve this, the hyperspectral data-cube is rearranged in a data matrix $D \in R^{(d \times n)}$ where n represents the number of pixels, and d is the spectral dimension. Subsequently, the data matrix is factorized as the product of k spectral endmembers represented by the matrix $W \in R^{(d \times k)}$ and the abundances or scores matrix $H \in R^{(k \times n)}$, which denotes the contribution of each endmember to a specific pixel. The process is summarized as follows:

$$D \approx WH \qquad (1)$$

* Corresponding author: alessia.dibenedetto@polimi.it

This optimization problem aims to minimize the difference between the original matrix and the obtained representation, influenced by the chosen matrix decomposition method. While Principal Component Analysis (PCA) remains widely recognized and utilized algorithm for matrix decomposition, newer [3] approaches have gained prominence in recent times, such as Non-negative Matrix Factorization (NMF) [3].

## 3 Fusion of hyperspectral datasets

Typically, hyperspectral data-cubes are handled separately, with each technique undergoing individual data evaluation. Comparison of the information obtained occurs only after the data reduction. However, the fusion of the hyperspectral datasets, when coupled to multivariate analysis, could provide a comprehensive description of the sample compounds. Through the fusion of different datasets, it is possible to harness the potential interactions between different imaging methods. Indeed, recent studies in CH field proved the effectiveness of this methodology [4,5], where data from different imaging modalities are combined in an early stage of the process, rather than during their interpretation. The fusion of hyperspectral data relies on algorithms originating from computer vision: for example, scale-invariant feature transform (SIFT) [4] is aimed to detect local features of the data and align the two data-frames by determine the transformation matrix needed to overlay one dataset onto another. This approach can be exploited for both vertical fusion, to merge multimodal datasets, as well as horizontal fusion, to enhance the quality and resolution of the hyperspectral images.

## 4 Example on real artwork


**Fig. 1. A)** Recto and **B)** verso of the manuscript's page. **C)** and **D)** RGB images of one of the areas of the verso in reflectance and transmittance configuration, respectively. **D)** and **E)** NMF analysis on transmittance and reflectance data, respectively, of the selected area. **F)** NMF analysis conducted on the fused data-frame of reflectance and transmittance configurations.

Here we present a real example of multivariate analysis applied to a fused hyperspectral dataset. The subject is a manuscript from Richthofen Collection in Tresoar (Leeuwarden) [6], which contains handwritten Frisian legal codes from early modern age. In one of these manuscript two pages were glued together and it could be seen in transmission illumination that on one of the glued pages text had been written. The aim was to obtain a readable representation of the text it with a non-invasive analysis. In Figure 1A and 1B the recto and verso of the glued page are shown. Transmission and reflectance measurements were conducted on both recto and verso. Each side was scanned in 6 sections and subsequentially fused together to enhance the spatial resolution. Figure 1E and 1F illustrate the outcome of a NMF analysis applied to the transmittance and reflectance configurations, respectively, in the form of the most legible image. In transmittance image, a blend of visible and hidden text is recognisable, while the reflectance image exclusively shows the visible text. Figure 1G present the NMF analysis performed on the fused dataset of both transmittance and reflectance of the verso in the same section. It is evident how the text becomes more visible and readable, as the features of both datasets complement each other, enabling a more precise discretization of the layers.

## 5 Conclusions

This contribution introduces a novel approach of multivariate analysis approach applied to fused hyperspectral datasets. The effectiveness of this methodology has been demonstrated using the example of a glued manuscript. Ongoing research involves testing this approach on the fusion of innovative combinations of datasets, such as merging Fluorescence Lifetime Imaging and Time-Gated Hyperspectral Imaging as well as Raman and optical fluorescence micro-mapping.

## References

[1] J. K. Delaney, K. A. Dooley, R. Radpour, and I. Kakoulli, Sci Rep **7**, (2017).

[2] R. Moreau, T. Calligaro, L. Pichon, B. Moignard, S. Hermon, and I. Reiche, X-Ray Spectrometry (2023).

[3] M. Alfeld, M. Wahabzad, C. Bauckhage, K. Kersting, G. Wellenreuther, P. Barriobero-Vila, G. Requena, U. Boesenberg, and G. Falkenberg, J Synchrotron Radiat **23**, 579 (2016).

[4] M. Alfeld, S. Pedetti, P. Martinez, and P. Walter, C R Phys **19**, 625 (2018).

[5] E. Catelli, Z. Li, G. Sciutto, P. Oliveri, S. Prati, M. Occhipinti, A. Tocchio, R. Alberti, T. Frizzi, C. Malegori, and R. Mazzeo, Anal Chim Acta **1239**, 340710 (2023).

[6] Book History Research Group Pastei, http://www.pastei.frl.