SF2A 2025
A. Siebert, K. Baillié, M. Béthermin, F. Cantalloube, E. Josselin, N. Lagarde, J. Malzac, J. Richard, L. Selliez, O. Venot (eds)

DOES DEEP PREVAIL FOR UNSUPERVISED CLUSTERING OF HYPERSPECTRAL IMAGES ?

H. J. Chambon[1] and D. Fraix-Burnety[1]

Abstract.
We explore unsupervised deep clustering and compare its results with the FisherEM algorithm on NGC1068 MUSE hyperspectral image. The deep approach is done by maximising a generalised mutual information with Maximum Mean Discrepancy and Unbiaised Sinkhorn Divergence. Classification results show the superiority of FisherEM to get large scale structure but the Unbiased Sinkhorn Divergence shows promising results to identify complex line morphologies.

Keywords: Galaxies, hyperspectral imaging, unsupervised classification

1 Introduction

With the advent of the MUSE spectrograph, wide fields hyperspectral images have become one of the most powerful tool to explore galaxies and the intergalactic medium. However, those images are complex, rendering the treatment of each individual spectrum an impossible or strenuous task. A common approach is spectral binning often associated with Voronoi tessellation but spatial proximity does not mean spectral similarity. Clustering has come as a natural solution to bin spectra with similar spectral features. Two different approach have been proposed : hierarchical clustering with the Euclidean distance (de Souza et al. 2025) and the FisherEM algorithm (Chambon & Fraix-Burnet 2024). However, both methods may struggle in front of the non-linearity brought by spectrum at different redshift. This is especially the case of the MUSE ultra deep fields were multiple galaxies of diverse and unknown redshift are observed. Supervised neural network are often needed to tackle non-linear problem, but no sufficient training sample is accessible today for the faint galaxies. The only solution is thus to explore unsupervised deep clustering.
Unsupervised deep clustering can tackle the clustering problem without prior assumption on the data distribution. The data distribution is modelled with a neural network. Given the the universal approximation theorems (Hornik et al. 1989) any continuous distributions were it linear or non-linear can then be estimated with a good network architecture. The traditional objective of deep clustering is to maximise the mutual information (Bridle, John S. et. al 1991). However the core distance of the mutual information, the Kullback-Leibler divergence, may not lead to satisfying results (Tschannen et al. 2019). Ohl et al. (2022) generalised the mutual information by incorporating geometrically aware core distances (GEMINI). We describe here the first results of unsupervised deep clustering relying on generalised mutual information for hyperspectral images and compare them to Chambon & Fraix-Burnet (2024).

2 NGC1068 MUSE Datacube

The NGC1068 MUSE Datacube is the same as in Chambon & Fraix-Burnet (2024), the data were retrieved from the ESO archive (ADP.2016-06-17T08:44:56.817). Due to poor quality the image border were removed. The spectra are normalised within the range [7600-8100Å] and corrected for internal motion of the galaxy with ALFA (Wesson 2016). Spectra with misleading redshift estimation were removed. Contrary to Chambon & Fraix-Burnet (2024) those spectra were further treated with a standardisation Z-score in the Maximum Mean Discrepancy (MMD) case and a normalisation to the total flux in the Unbiaised Sinkhorn Divergence (USD) case.

[1] Univ. Grenoble Alpes, CNRS, IPAG, Grenoble, France

© Société Française d'Astronomie et d'Astrophysique (SF2A) 2025

3 Method

3.1 FisherEM

FisherEM (Bouveyron & Brunet 2012) is a clustering algorithm based on parsimonious Gaussian Mixture models in a discriminative subspace. To solve high-dimensional drawbacks, it assume that the data lives in a subspace of lower dimension. The optimisation of the mixture parameters is done with the expectation maximisation (EM) algorithm (Moon 1996). The subspace is optimised with an additional step based on the Fisher criterion that reduces the internal covariance of the classes and increases the in between covariance of the classes (Fisher 1936). The number of classes and the choice of the gaussian model is done with the integrated complete likelihood criteria (Biernacki et al. 2000). FisherEM has been used for years to deal with astronomical spectra (Fraix-Burnet et al. 2021) and has been found relevant in the hyperspectral images context (for more details see Chambon & Fraix-Burnet 2024).

3.2 Unsupervised deep clustering with GEMINI

3.2.1 GEMINI implementation and parameters

GEMINI is implemented in the python GemClus library. The default architecture is a two layer neural network architecture with ReLu activation function. Its hidden layer has 20 neurons, due to the high dimensional nature of the data we increased the size of this hidden layer to 50 neurons.
GemClus offers two paradigm of optimisation : each cluster distribution can be compared either to the whole distribution (OVA) or to each cluster distribution (OVO). OvO needs quadratic number of GEMINI estimates and OvA a linear one resulting in different time complexity. OvO leads to finer cluster. Contrary to FisherEM, GEMINI is able to choose the right number of class but needs an upper bound on its number.

3.2.2 GEMINI MMD

MMD is a distance measure between two distributions from their sample represented in a reproducing kernel hilbert space (RKHS). To estimate MMD with a monte carlo method GEMINI uses the kernel trick and the Bayes theorem (Ohl et al. 2022) leading to the GEMINI Eq.(3.1) for the OvO case and Eq.(3.2) for the OVA one.

$$E_{y_a,y_b\sim p_\theta(y)}\left[E_{x_a,x_b\sim p_{data}(x)}\left[k(x_a,x_b)\left(\frac{p_\theta(y_a|x_a)p_\theta(y_a|x_b)}{p\theta(y_a^2)}\right)+\frac{p_\theta(y_b|x_a)p_\theta(y_b|x_b)}{p\theta(y_b^2)}-2\frac{p_\theta(y_a|x_a)p_\theta(y_a|x_b)}{p_\theta(y_a)p_\theta(y_b)}\right]^{\frac{1}{2}}\right] \tag{3.1}$$

$$E_{p_\theta(y)}\left[E_{x_a,x_b\sim p_{data}(x)}\left[k(x_a,x_b)\left(\frac{p_\theta(y_a|x_a)p_\theta(y_a|x_b)}{p\theta(y_a^2)}\right)+1-2\frac{p_\theta(y|x_a)}{p_\theta(y)}\right]^{\frac{1}{2}}\right] \tag{3.2}$$

3.2.3 Unbiaised Sinkhorn Divergence GEMINI

The Wasserstein Distance is an optimal transport distance between distributions defined as a minimisation problem 3.3 with $\Gamma(p,q)$ the set of all couplings between p and q and c a distance function. As clustering samples are finite and not complete distribution, Ohl et al. (2022) approximates the distribution with weighted sum of dirac (see their eq 11).

$$W_c(p,q)=\left(\inf_{\gamma\in\Gamma(p,q)}E_{x,z\sim\gamma(x,z)}[c(x,z)]\right) \tag{3.3}$$

The computation of the Wasserstein Distance in the high-dimensional context with high number of samples is computationally intensive both in terms of time complexity ($O(n^3\log(n))$ and spatial complexity in $O(n^2)$ due to a cost matrix (Flamary et al. 2021). In contrast, the Unbiased Sinkhorn divergence (USD) approach from Feydy et al. (2018) has a time complexity of less than $O(n^2)$ and a spatial complexity of $O(n)$. With the ϵ parameter the USD interpolates between the true Wasserstein distance ($\epsilon = 0$) and MMD ($\epsilon = +\infty$). We made a trade off with $\epsilon = 0.25$. We chose the online implementation over the multiscale one due to poor result

with Kmeans in the high dimensional context and escape an additional parameter with the number of Kmeans classes. This led us to two new GEMINI the USD OVO with Eq.(3.4) and USD OVA with Eq.(3.5).

$$E_{y_a,y_b\sim p_\theta(y)}\left[USD_c\left(\sum_{i=1}^{N} m_i^{y_a}\delta_{x_i}, \sum_{i=1}^{N} m_i^{y_b}\delta_{x_i}\right)\right] \tag{3.4}$$

$$E_{p_\theta(y)}\left[USD_c\left(\sum_{i=1}^{N} m_i^{y}\delta_{x_i}, \sum_{i=1}^{N} \frac{1}{N}\delta_{x_i}\right)\right] \tag{3.5}$$

4 Comparison between FisherEM and unsupervised deep clustering

4.1 FisherEM

The clustering of NGC1068 with FisherEM led to an optimum of 16 classes. The map of classes Fig. 1 depicts the complex morphology of this spiral galaxy (ring of star-formation, spiral arms, and AGN). The classes 10,11 and 12 are characteristic of the influence of the jet with high OIII ionisation. Classes 7, 13, 14 and 15 are of the HII type in the BPT diagram. An additional sub-classification of the class 15 was made to disentangle the spaxels near the AGN and those in the ring (for more details see Chambon & Fraix-Burnet 2024). We take this clustering as a reference.

4.2 GEMINI MMD

The clustering of non-standardised NGC1068 spectra with GEMINI MMD failed to encompass the diversity inside the galaxy with very few classes typical of distinct physical origin. The clustering of normalised and standardised Z-score spectra with an upper bound of 21 classes recovers the broad morphology of NGC1068 (see fig. 2) yet it fails to distinguish HII regions corresponding to classes 7, 13, 14 and 15 of FisherEM, and regions characteristic of the jet influence corresponding to classes 10, 11 and 12 of FisherEM. Still, it separates the nucleus from the other classes. Overall, the GEMINI MMD clustering is less informative. This could be due to its clustering objective similar to Kmeans (Ohl et al. 2022).

4.3 GEMINI Unbiaised Sinkhorn Divergence

The clustering of NGC1068 with USD in the OvO paradigm with an upper bound of 11 classes and 100 iterations led to 7 classes (see fig. 2). This classification separates the AGN (classes 3 to 7), the HII region (class 2) and the Interstellar medium (class 1). Interestingly, the AGN classes show different line morphologies that were not highlighted by FisherEM even with sub-classification. This exhibits better sensibility to kinematics.
However, USD fails to distinguish analogs to classes 10, 11 and 12 from FisherEM and its description of ISM is lacking. One key point to note is that 100 iterations are not enough to reach classification convergence. Thus, more iterations are needed to conclude on any superiority of USD. Yet, USD is much more time consuming that FisherEM.

5 Conclusions

Unsupervised classification of spaxels can be tackled with both linear and non-linear methods. On the NGC1068 MUSE hyperspectral image, the linear FisherEM method outperforms the GEMINI MMD and renders a more complete picture than GEMINI USD. However, GEMINI USD reaches greater precision regarding the complex kinematics of the AGN without sub-classification.

Thank you to the organizers of the SF2A 2025!
All (or most of) the computations presented in this paper were performed using the GRICAD infrastructure (https://gricad.univ-grenoble-alpes.fr), which is supported by Grenoble research communities.

References

Biernacki, C., Celeux, G., & Govaert, G. 2000, IEEE Transactions on Pattern Analysis and Machine Intelligence, 22, 719
Bouveyron, C. & Brunet, C. 2012, Statistics and Computing, 22, 301, number: 1

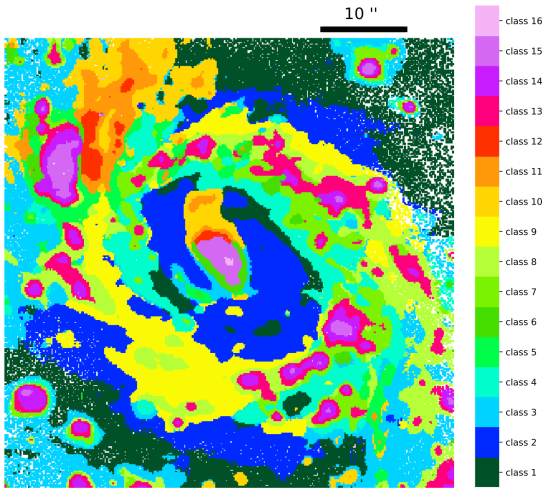

Fig. 1. Reference classification map of NGC1068 with FisherEM

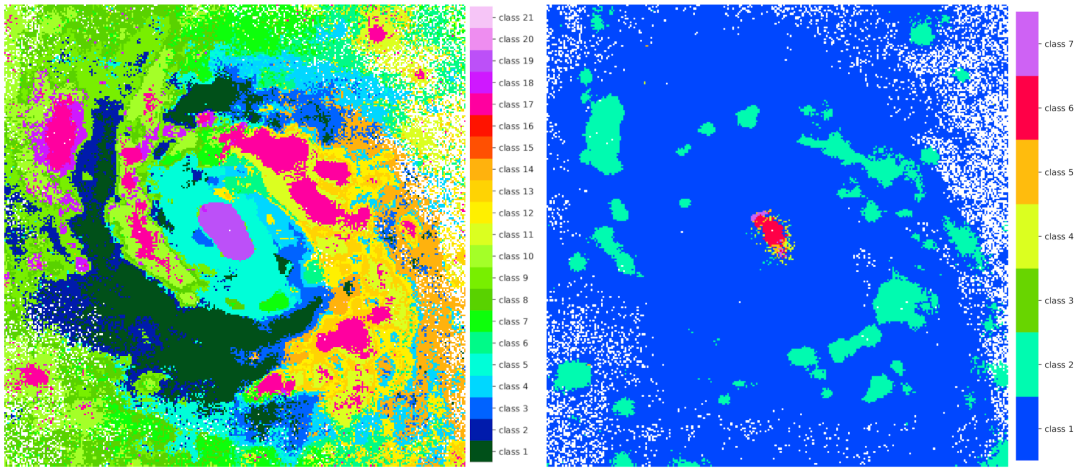

Fig. 2. Left: Classification map of NGC1068 with MMD OvO and 21 asked classes with 1000 iterations. **Right:** Classification map of NGC1068 with USD OvO and 11 asked classes with 100 iterations.

Bridle, John S. et. al. 1991, in Proceedings of the 5th International Conference on Neural Information Processing Systems, NIPS'91 (Denver, Colorado), 1096–1101

Chambon, H. J. & Fraix-Burnet, D. 2024, Astronomy & Astrophysics, 688, A19

de Souza, R. S., Dahmer-Hahn, L. G., Shen, S., et al. 2025, Monthly Notices of the Royal Astronomical Society, 539, 3166

Feydy, J., Séjourné, T., Vialard, F.-X., et al. 2018, Interpolating between Optimal Transport and MMD using Sinkhorn Divergences, version Number: 1

Fisher, R. A. 1936, Annals of Eugenics, 7, 179

Flamary, R., Courty, N., Gramfort, A., et al. 2021, Journal of Machine Learning Research, 22, 1

Fraix-Burnet, D., Bouveyron, C., & Moultaka, J. 2021, Astronomy & Astrophysics, 649, A53

Hornik, K., Stinchcombe, M., & White, H. 1989, Neural Networks, 2, 359

Moon, T. 1996, IEEE Signal Processing Magazine, 13, 47

Ohl, L., Mattei, P.-A., Bouveyron, C., et al. 2022, in Advances in Neural Information Processing Systems, ed. S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, & A. Oh, Vol. 35 (Curran Associates, Inc.), 3377–3390

Tschannen, M., Djolonga, J., Rubenstein, P. K., Gelly, S., & Lucic, M. 2019, On Mutual Information Maximization for Representation Learning, version Number: 2

Wesson, R. 2016, Monthly Notices of the Royal Astronomical Society, 456, 3774, aDS Bibcode: 2016MNRAS.456.3774W