Exploring the Universe with Artificial Intelligence (UniversAI)
Proceedings IAU Symposium No. 397, 2026
I. Liodakis & V. Efthymiou, eds.
doi:10.1017/S1743921325104535

Potential of Feature Extraction Method Using Kernel PCA and Kernel SHAP in Astronomy

Hiroma Okubo[1], Tsutomu T. Takeuchi[2,3], Shotaro Akaho[4,5], Yasuhiko Igarashi[6,7], Nario Kuno[1,8], Toshiki Saito[9], Nanase Harada[10,11], Akio Taniguchi[12], Shuro Takano[13] and Taku Nakajima[14]

[1]Division of Physics, Faculty of Pure and Applied Sciences, University of Tsukuba, Tsukuba, Ibaraki 305-8571, Japan
email: s2430049@u.tsukuba.ac.jp

[2]Division of Particle and Astrophysical Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8602, Japan

[3]The Research Center for Statistical Machine Learning, The Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan

[4]Human Informatics and Interaction Research Institute, The National Institute of Advanced Industrial Science and Technology, Tsukuba, Japan

[5]Center for Training Professors in Statistics, The Institute of Statistical Mathematics, Tachikawa, Tokyo, Japan

[6]Institute of Engineering, Information and Systems, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki, 3058573, Japan

[7]Photoemission Group, National Institute for Materials Science, 3-13 Sakura, Tsukuba, Ibaraki, 3050003, Japan

[8]Tomonaga Center for the History of the Universe, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8571, Japan

[9]Faculty of Global Interdisciplinary Science and Innovation, Shizuoka University, 836 Ohya, Suruga-ku, Shizuoka 422-8529, Japan

[10]National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan

[11]Astronomical Science Program, Graduate Institute for Advanced Studies, SOKENDAI, 2-21-1 Osawa, Mitaka, Tokyo 181-1855, Japan

[12]Kitami Institute of Technology, 165 Koen-cho, Kitami, Hokkaido 090-8507, Japan

[13]Department of Physics, General Studies, College of Engineering, Nihon University, Tamura-machi, Koriyama, Fukushima 963-8642, Japan

[14]Institute for Space-Earth Environmental Research, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8601, Japan

Abstract. We present a new framework for multi-line analysis that combines kernel principal component analysis (kernel PCA) and kernel shapley additive explanations (kernel SHAP). Applied to ALMA Band 3 data of NGC 1068, kernel PCA reveals features—including the AGN-driven outflow —not clearly captured by PCA, while kernel SHAP quantifies each line's contribution. These results are consistent with previous studies and confirmed by a two-sample t-test. Moreover, kernel PCA scores correlate with LTE-derived column densities, highlighting the robustness of this framework.

Keywords. methods: data analysis, galaxies: ISM, galaxies: active, ISM: molecules

© The Author(s), 2026. Published by Cambridge University Press on behalf of The International Astronomical Union.

https://doi.org/10.1017/S1743921325104535 Published online by Cambridge University Press

1. Introduction

Multi-line analysis are a powerful tool to probe the physical and chemical conditions of the interstellar medium in galaxies (e.g., Takano et al. 2014; Harada et al. 2019). Linear feature extractions have been used for multi-line analysis (e.g., Meier & Turner 2005; Kishikawa et al. 2025). But these methods cannot fully capture the non-linear variations arising from radiative transfer, chemistry, and excitation (Okubo et al. 2025). To overcome this limitation, we propose a new framework that combines kernel PCA with kernel SHAP. We explain the potential of this method by presenting the data and methodology in Section 2, the application of our approach in Section 3, and the discussion of the results in Section 4.

2. Data and Methods

We use archival ALMA Band 3 integrated intensity maps of NGC 1068 with a spatial resolution of about 150 pc. In total, we use 13 molecular line data: C_2H (N=1–0), HCN (J=1–0), HCO^+ (J=1–0), HNC (J=1–0), HC_3N (J=10–9), N_2H^+ (J=1–0), CH_3OH ($J=2_K$–1_K), CS (J=2–1), $C^{18}O$ (J=1–0), ^{13}CO (J=1–0), CN ($N=1_{3/2}$–$0_{1/2}$), CN ($N=1_{1/2}$–$0_{1/2}$), and CO (J=1–0). These maps are produced using standard calibration and imaging with the `PHANGS_ALMA` pipeline (Leroy et al. 2021) and then standardized.

Kernel PCA extends conventional PCA by mapping the data into a higher-dimensional space, allowing it to capture non-linear relationships. We adopt the radial basis function (RBF) kernel with a kernel width of 0.01, selected after testing values of 0.1, 0.01, 0.001, ... geometrically, as it yielded results most consistent with astronomical interpretations.

Kernel SHAP approximates the SHAP method, assigning each molecular line an importance value for its contribution to a given principal component. As a model-agnostic approach, it can be applied directly to kernel PCA results, providing quantitative interpretation of each line's role. By doing so, Kernel SHAP can overcome previous difficulties in applying kernel PCA to multi-line analyses, making such applications feasible.

3. Results

The first three kernel principal components (KPCs) account for over 90% of the total variance, with KPC1, KPC2, and KPC3 contributing 68%, 20%, and 5%, respectively. Figure 1 shows the spatial distributions of KPC1–3, while Figure 2 presents the corresponding kernel SHAP values for each molecular line.

The KPC1 and KPC2 maps are consistent with the PCA results reported by Okubo et al. (2025), and their associated kernel SHAP distributions exhibit trends similar to the PCA coefficients. Although kernel PCA can, in principle, lose some of the original data's information, these results demonstrate that the method preserves physically meaningful structures. Detailed interpretations of the physical implications are provided in Okubo et al. (2025).

For KPC3, high positive scores appear to be concentrated around the AGN, while low positive scores are found near the spiral arms. Negative scores look like to be elongated in the central region. The kernel SHAP analysis shows that CO isotopologues generally contribute positively, while HC_3N and N_2H^+ exhibit strong positive contributions localized to the AGN grid. Conversely, CN and C_2H have strong negative contributions, followed by HCN and HNC with moderately negative contributions.

4. Discussion

We focus here on the newly extracted KPC3 map. The low positive scores are distributed along the spiral arms, with kernel SHAP indicating a dominant contribution

https://doi.org/10.1017/S1743921325104535 Published online by Cambridge University Press

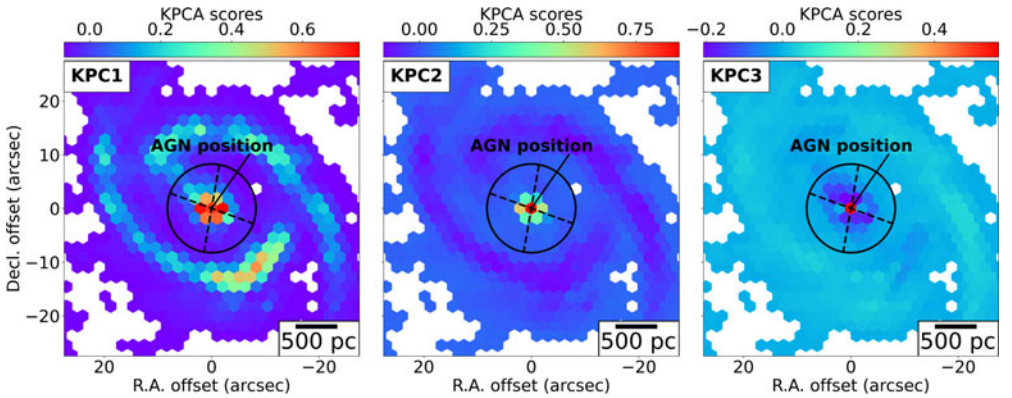


Figure 1. Spatial distributions of the first three kernel principal components (KPC1–3). The AGN position is indicated by Saito et al. (2022). Color bars represent the KPC scores.

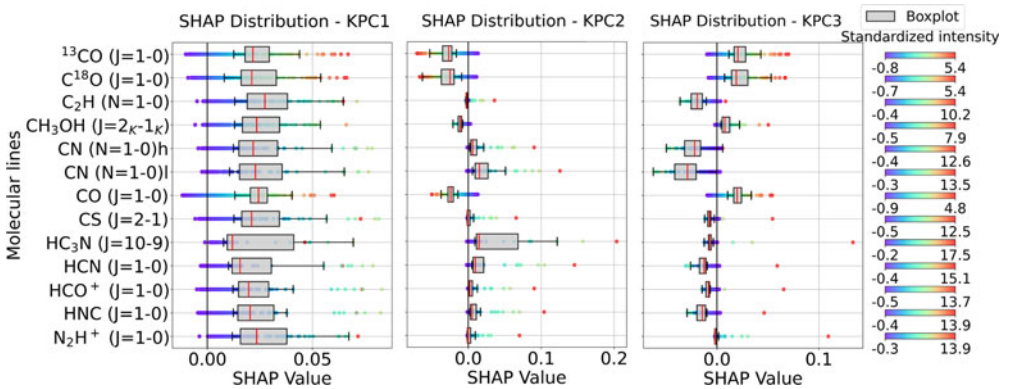


Figure 2. Kernel SHAP value distributions for each molecular line corresponding to KPC1–3. Boxplots show the distribution of SHAP values, with color representing the standardized integrated intensity.

from CO isotopologues. This suggests that these scores trace low-density gas in the arms. In contrast, the high positive scores are concentrated in the AGN region, with strong contributions from HC_3N, HCN, N_2H^+, CS, and HCO^+. These molecules are known tracers of dense gas, and their distribution implies that KPC3 reflects both the dense environments in the galactic center and the chemically rich conditions influenced by the AGN (e.g., Harada et al. 2010, 2013; Kauffmann et al. 2017; Nakajima et al. 2018; Butterworth et al. 2022).

We next consider the negative scores. These regions are dominated by contributions from CN and C_2H, followed by HNC, HCN, CS and HCO^+. This distribution suggests that the negative scores may trace outflows or active star-forming regions (e.g., Aalto et al. 2012; García-Burillo et al. 2017; Saito et al. 2022). Notably, molecules such as HC_3N and N_2H^+, which trace dense gas, are absent in these regions. Furthermore, a two-sample t-test at the 5% significance level indicates that the negative-score regions align with the known outflow axis, supporting the interpretation that KPC3 negative scores capture outflow-related structures.

To further investigate the meaning of KPC3, we performed an LTE analysis to estimate molecular column densities within the central 1 kpc, and correlated them with negative KPC3 scores (< -0.03) (Goldsmith & Langer 1999). Only lines with S/N $\geq$ 4 and at

https://doi.org/10.1017/S1743921325104535 Published online by Cambridge University Press

Table 1. Correlation coefficients between molecular column densities and KPC3 scores (< -0.03) in the central 1 kpc.

Molecular line	^{13}CO	C_2H	CN	CS
Correlation	0.069±0.027	-0.628±0.023	-0.844±0.017	-0.782±0.006
	HCN	HNC	HCO^+	
	-0.816±0.004	-0.842±0.003	-0.800±0.007	

least 10 valid spatial grids were used. The results, summarized in Table 1, show the correlations between KPC3 and molecular column densities together with their uncertainties, which were estimated via Monte Carlo sampling. Molecules identified by kernel SHAP as contributing to the negative scores (CN, C_2H, HCN, HNC, HCO^+, CS) exhibit strong negative correlations with the KPC3 scores, whereas ^{13}CO, not identified as contributing to the negative scores, shows no significant correlation. This consistency demonstrates that the kernel SHAP results indeed extract physically meaningful information from the molecular line data.

5. Summary

Kernel PCA extracted features not captured by conventional PCA. We demonstrate, based on previous studies and additional analyses, that the features interpreted by kernel SHAP correspond to physically meaningful structures. These results highlight the effectiveness of our proposed framework.

Acknowledgements

We thank the organizers of IAU Symposium 397 "UniversAI" for the opportunity to present this work. This paper makes use of the following ALMA data: ADS/JAO.ALMA#2011.0.00061.S, ADS/JAO.ALMA#2012.1.00657.S, ADS/JAO.ALMA#2013.1.00060.S, ADS/JAO.ALMA#2013.1.00279.S, ADS/JAO.ALMA#2015.1.00960.S, ADS/JAO.ALMA#2017.1.00586.S, ADS/JAO.ALMA#2018.1.01506.S, ADS/JAO.ALMA#2018.1.01684.S, and ADS/JAO.ALMA#2019.1.00130.S. ALMA is a partnership of ESO (representing its member states), NSF (USA), and NINS (Japan), together with NRC (Canada), MOST and ASIAA (Taiwan), and KASI (Republic of Korea), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, AUI/NRAO, and NAOJ. This research was supported by a grant from the Hayakawa Satio Fund awarded by the Astronomical Society of Japan. This work was supported by JST SPRING, Grant Number JPMJSP2124. We acknowledge the use of ChatGPT (OpenAI, San Francisco, CA, USA) for language polishing assistance during manuscript preparation. All responsibility for the content remains with the authors.

References

Aalto, S., Garcia-Burillo, S., Muller, S., et al. 2012, Astronomy & Astrophysics, 537, A44, doi: 10.1051/0004-6361/201117919

Butterworth, J., Holdship, J., Viti, S., & García-Burillo, S. 2022, Astronomy & Astrophysics, 667, A131, doi: 10.1051/0004-6361/202244563

García-Burillo, S., Viti, S., Combes, F., et al. 2017, Astronomy & Astrophysics, 608, A56, doi: 10.1051/0004-6361/201731862

Goldsmith, P. F., & Langer, W. D. 1999, Astrophysical Journal, 517, 209, doi: 10.1086/307195

Harada, N., Herbst, E., & Wakelam, V. 2010, Astrophysical Journal, 721, 1570, doi: 10.1088/0004-637X/721/2/1570

Harada, N., Sakamoto, K., Martín, S., et al. 2019, Astrophysical Journal, 884, 100, doi: 10.3847/1538-4357/ab41ff

Harada, N., Thompson, T. A., & Herbst, E. 2013, Astrophysical Journal, 765, 108, doi: 10.1088/0004-637X/765/2/108

https://doi.org/10.1017/S1743921325104535 Published online by Cambridge University Press

Kauffmann, J., Goldsmith, P. F., Melnick, G., et al. 2017, Astronomy & Astrophysics, 605, L5, doi: 10.1051/0004-6361/201731123

Kishikawa, R., Harada, N., Saito, T., et al. 2025, Publications of the Astronomical Society of Japan, 77, 1, doi: 10.1093/pasj/psae095

Leroy, A. K., Hughes, A., Liu, D., et al. 2021, Astrophysical Journal Supplement Series, 255, 19, doi: 10.3847/1538-4365/abec80

Meier, D. S., & Turner, J. L. 2005, Astrophysical Journal, 618, 259, doi: 10.1086/426499

Nakajima, T., Takano, S., Kohno, K., Harada, N., & Herbst, E. 2018, Publications of the Astronomical Society of Japan, 70, 7, doi: 10.1093/pasj/psx153

Okubo, H., Saito, T., Takano, S., et al. 2025, arXiv e-prints, arXiv:2506.15999, doi: 10.48550/arXiv.2506.15999

Saito, T., Takano, S., Harada, N., et al. 2022, Astrophysical Journal, 935, 155, doi: 10.3847/1538-4357/ac80ff

Takano, S., Nakajima, T., Kohno, K., et al. 2014, Publications of the Astronomical Society of Japan, 66, 75, doi: 10.1093/pasj/psu052

https://doi.org/10.1017/S1743921325104535 Published online by Cambridge University Press