ML Classification of Synthetic Spectra:
Be stars and N_2H^+

Fredy Orjuela[1], *Beatriz Sabogal*[1]
[1]*Departamento de Física, Universidad de los Andes, Bogotá, Colombia*
Email: fa.orjuela@uniandes.edu.co

Abstract

We present a machine-learning and Bayesian framework based on a neural-network emulator of HDUST/BeAtlas spectra, originally developed for classical Be stars and later adapted to model N_2H^+(1–0) emission in the protocluster G327.29 observed by the ALMA-IMF survey. The method is designed to characterize the dense and kinematically complex environments typical of the Central Molecular Zone (CMZ). Uncertainties from the emulator are incorporated directly into the statistical inference, while multiple sampling approaches are used to explore the posterior distributions and identify the best-performing model. Applied to the Be star γCas, the framework successfully recovers the disk density structure and surface density with high precision, with values of $n = 2.14^{+0.09}_{-0.06}$ and $\Sigma_0 = 0.46 \pm 0.03$ g cm^{-2}, respectively.

Introduction

Be stars (L-band)

- **Context:** Classical Be stars host viscous decretion disks (VDD) directly associated with the emission lines in their spectra.
- **Problem:** We have L-band data with a low resolution of $R \sim 500$, which causes strong degeneration in the VDD parameters: mass M, base surface density Σ_0, density slope n, and inclination $\cos i$.

N_2H^+

- **Context:** Star formation in the Central Molecular Zone (CMZ) occurs under extreme turbulence and magnetic fields; N_2H^+ is a crucial tracer for these high-pressure environments.
- **Problem:** Modeling the seven-component hyperfine structure on a pixel-by-pixel basis across the ALMA-IMF cube becomes computationally prohibitive.

Objectives

- **Be stars:** Build a neural emulator to accelerate Bayesian inference, breaking degeneracies by combining L-band diagnostics [4].
- **N_2H^+:** Extend the same framework to N_2H^+(1–0) emission, fitting one-, two-, or three-component velocity scenarios across ALMA-IMF protocluster data.

Methodology

Pipeline & Forward Models

1. HDUST/BeAtlas [1] models are used for classical Be stars, and LOC [2] radiative-transfer simulations for the N_2H^+(1–0) emission under one-, two-, and three-component scenarios.
2. A neural-network emulator is trained for each observable using signed-log targets, with performance evaluated through parity plots and residual distributions.

Bayesian Framework

3. **Likelihood:** Gaussian, incorporating emulator uncertainty: $\sigma_{\rm tot} = \sqrt{\sigma^2_{\rm obs} + \sigma^2_{\rm sys} + \sigma^2_{\rm NN}}$.
4. **Inference:** Multi-engine approach combining several sampling methods and variational-inference techniques to explore the posterior distributions.

Outputs (Be): WAIC-based model selection and MCMC posterior constraints for the viscous decretion disk parameters of γCas (Σ_0, n) [5].

Outputs (N_2H^+): Multi-engine Bayesian inference provides constraints on $V_{\rm LSR}$, excitation temperature, and optical depth, with transfer learning adapted to the noise properties of ALMA-IMF data [3].

Results and Analyses

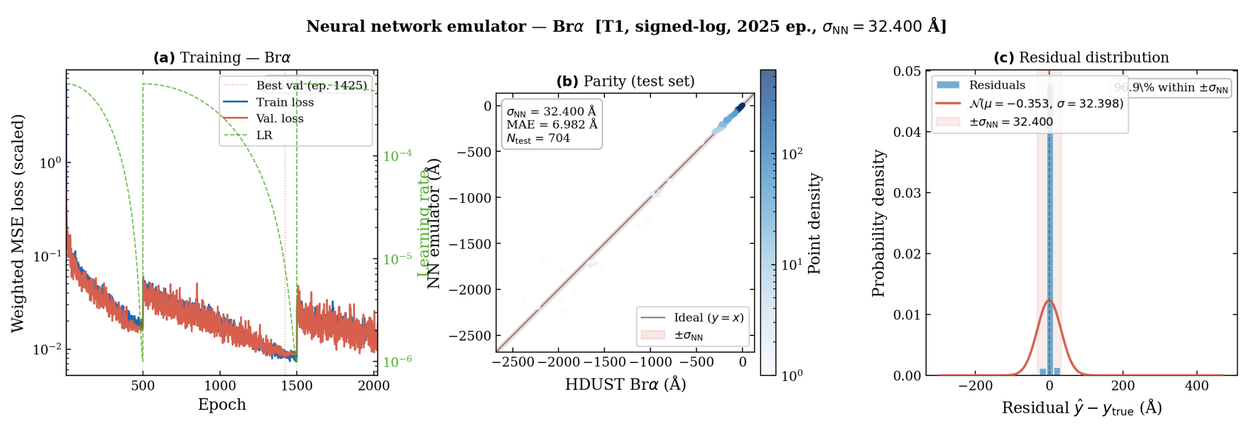

Figure A: Neural-network emulator validation for the Brα spectral line. **(a)** Evolution of the weighted Mean Squared Error (MSE) loss for the training (blue) and validation (red) sets across 2025 epochs; the dashed green line indicates the cyclical learning-rate schedule, with the best model state captured at epoch 1425 (dotted vertical line). **(b)** Parity plot comparing the emulator predictions against synthetic HDUST flux values for the test set ($N_{\rm test} = 704$); the high linearity and low Mean Absolute Error (MAE $\approx$ 6.98 Å) demonstrate the emulator's reliability across the VDD parameter space. **(c)** Probability density distribution of the residuals ($\hat{y} - y_{\rm true}$), fitted with a Gaussian ($\mu \approx -0.35$ Å, $\sigma \approx 32.40$ Å); 96.9% of the predictions fall within $\pm\sigma_{\rm NN}$, confirming the absence of significant systematic biases.

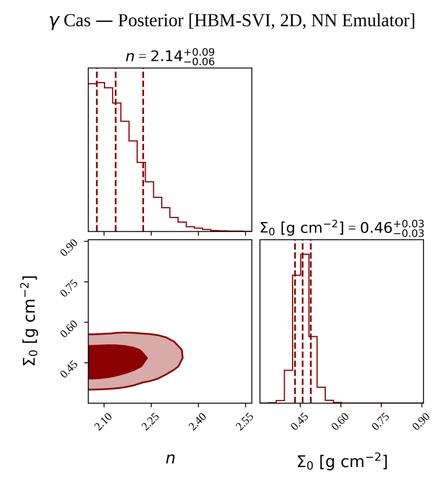

Figure B: Posterior distributions for γCas. Joint and marginalised probability density functions (PDFs) for the disk power-law exponent (n) and base surface density (Σ_0). The 2D contours represent the 1σ, 2σ, and 3σ confidence levels. The narrow, unimodal univariate distributions indicate successful parameter identifiability and convergence.

Table 1 — Statistical Results for γCas

Quantity	This Work	Vieira+17	Stat.
n	$2.14^{+0.09}_{-0.06}$	3.0–3.5	1σ
Σ_0 [g cm^{-2}]	0.46 ± 0.03	—	—
$\log \rho_0$	-10.96 ± 0.05	-10.6 ± 0.3	1.2σ
M [$M_\odot$]	16.2 ± 1.0	25 ± 9	1σ

Posterior estimates compared to the values reported by Vieira et al. (2017) [4].

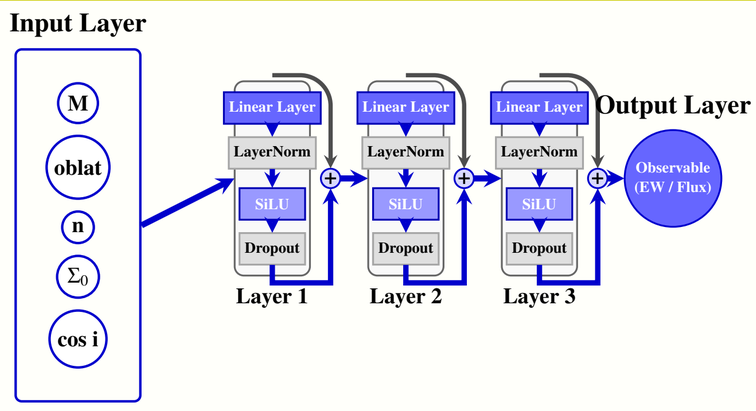

Figure C: BeAtlas Deep Emulator Topology. A ResNet architecture mapping the 5D parameter space ($\vec{\theta} \in \mathbb{R}^5$) to stellar observables. **Pipeline:** integrated residual hierarchy using SiLU activation and LayerNorm for optimised convergence. Uncertainty $\sigma_{\rm NN}$ is quantified through a non-parametric head for Bayesian consistency.

Future Work

- **Scalability to the CMZ:** The Bayesian–ML pipeline is uniquely suited to decompose the complex, multi-component velocity fields of the Galactic Center. We aim to apply this framework to ALMA-IMF datasets [3] in the CMZ to map star-formation efficiency where conditions are unlike any other place in the Milky Way.

References

[1] Carciofi & Bjorkman (2006), *ApJ* 639, 1081 — HDUST.
[2] Juvela (2020), *A&A* 644, A151 — LOC.
[3] Motte *et al.* (2022), *A&A* 662, A8 — ALMA-IMF.
[4] Vieira *et al.* (2017), *MNRAS* 464, 3071.
[5] Millán-Valderrama *et al.* (2025), *PASP* 137, 044204.

Acknowledgements

Thanks to the organizers of the ***Traversing the Galactic Center in Space and Time*** symposium; **IAU** travel grant; and Vicerrectoría de Investigaciones at Universidad de los Andes, grant INV-2026-205-3268. We also thank the **ALMA-IMF** collaboration, particularly Prof. Amelia Stutz and Rodrigo Álvarez, for the reduced and processed data, which under authorization will be used for subsequent training of the framework.