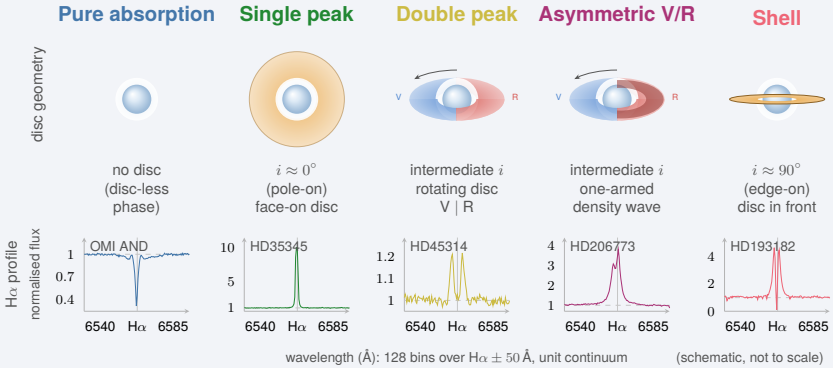
BESS-Bench

Benchmarking Spectral Representations for Be-Star Variability

Matthieu Le Lain[1] · Sébastien Lefèvre[1,2]

[1]IRISA, UMR CNRS 6074, Université Bretagne Sud, Vannes, France
[2]UiT – The Arctic University of Norway, Tromsø, Norway

NeurIPS 2026 · Evaluations & Datasets Track · Poster 139851



- Rapidly rotating B stars build and lose a **decretion disc**: the Hα line varies on **days-to-decades** timescales.
- The *same* star transits between these morphologies: a **multi-epoch** problem.

The data gap

1 epoch per star (SDSS, APOGEE, LAMOST)

vs

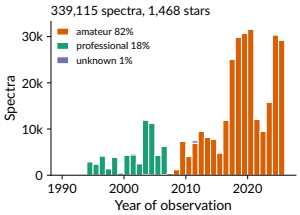
339 k multi-epoch BeSS spectra **never curated for ML**

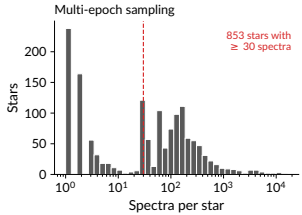
339,115
optical spectra
1990–2025

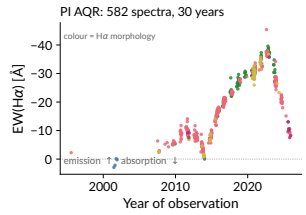
1,468
Be stars
multi-epoch

81.7 %
amateur
observations (BeSS)

26,858
scored Hα spectra
1,428 stars





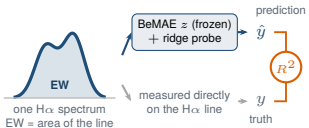


Scored Hα slice
128 bins over H$\alpha \pm 50$ Å,
6 analytic line features

No quality cut on the corpus
SNR is a covariate; SNR ≥ 10
gate only on the scored slice

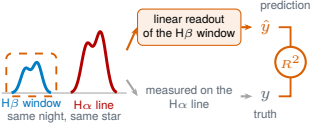
Frozen star-level split
1,024 / 219 / 225 stars,
Croissant, DOI

SpecProbe



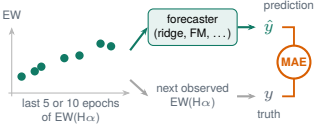
y = EW, peak, FWHM, Δv, V/R, depth

LineTransfer



y = EW, peak, FWHM, depth of Hα

EWForecast



MAE / $\text{MAE}_{\text{persist}}$ (< 1 wins)

One protocol for all

Stars, not spectra, are split

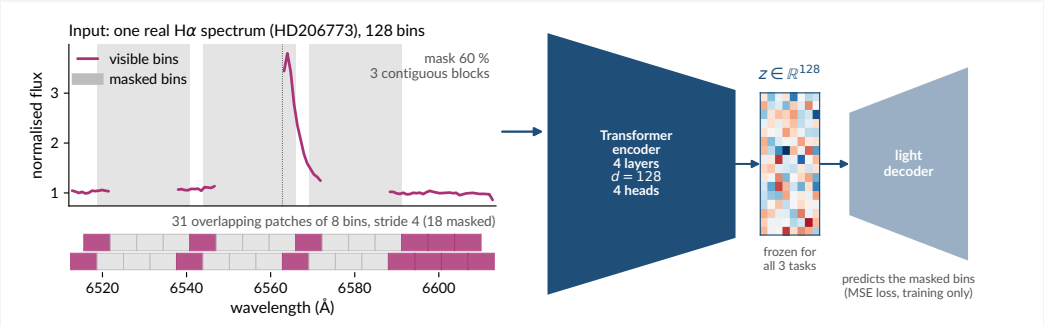
5-fold GroupKFold; temporal cutoff 2021-01-01 for forecasting.

3 seeds, bootstrap CI_{95}

10,000 resamples; one aggregated ratio ρ against PCA(10).

Who competes

BeMAE, PCA + ridge, raw flux, persistence, zero-shot foundation models.



One real spectrum

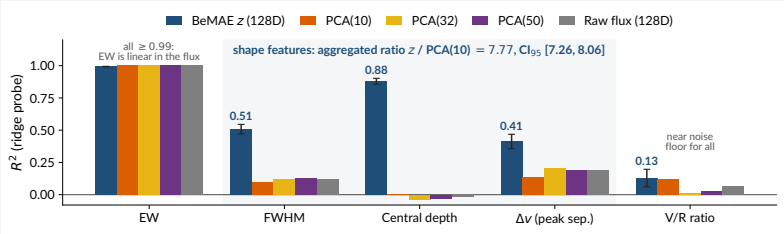
HD206773, 128 bins $\rightarrow$ 31 overlapping patches, 60 % masked in 3 blocks.

A small encoder

4-layer Transformer, $d = 128$, **795 k parameters**; self-supervised, 3 seeds released.

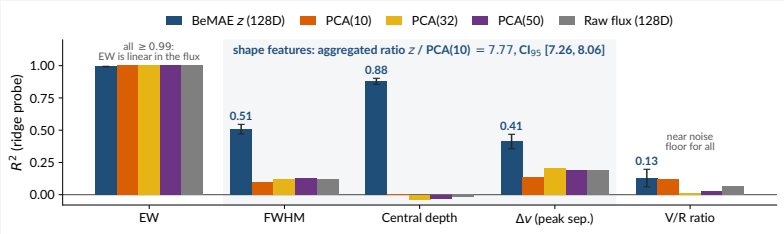
One frozen embedding

$z \in \mathbb{R}^{128}$, used as is by every task; its first PC separates absorption from emission.



R^2 of a ridge probe on frozen representations, star-grouped 5-fold CV; mean over 3 encoder seeds, whiskers $\pm$ std.

- **EW** ($R^2 \geq 0.99$ for everyone) is linear in the flux: no representation needed.



R^2 of a ridge probe on frozen representations, star-grouped 5-fold CV; mean over 3 encoder seeds, whiskers $\pm$ std.

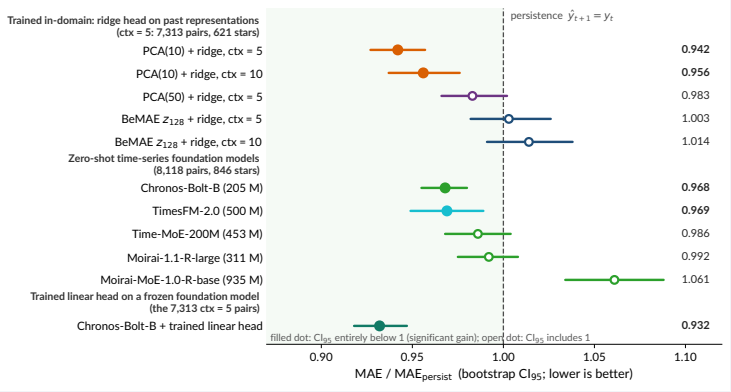
- **EW** ($R^2 \geq 0.99$ for everyone) is linear in the flux: no representation needed.
- **Line shape is not**: BeMAE reaches $R^2 = 0.51$ / 0.88 / 0.41 on FWHM / central depth / Δv where every linear readout stays < 0.21.

Aggregated ratio

$$\rho = 7.77$$

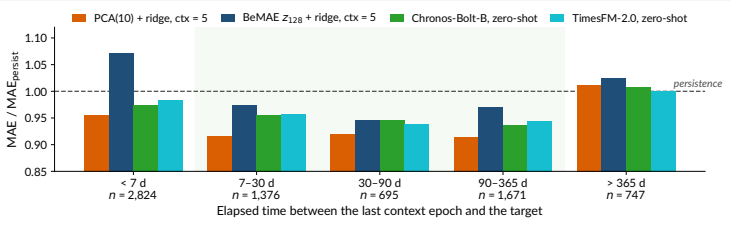
CI_{95} [7.26, 8.06], 3 seeds

The raw flux (100 % of the variance) fails too: **linear compressibility $\neq$ physical informativeness**.



One-step EW(Hα) forecasting under the temporal cutoff. Ratio < 1: better than repeating the last observation. Filled dot: CI_{95} entirely below 1.

- Best in-domain: **PCA(10) + ridge, 0.942** ($-5.8\,\%$ MAE).
- **BeMAE** 1.003: EW is near-linear, PCA(10) recovers it exactly.
- Nine zero-shot foundation checkpoints: $\leq 3.2\,\%$ gain; a trained linear head on frozen Chronos-Bolt-B: **0.932**.



Horizon dependence: baselines beat persistence on 7–365 d; beyond one year every ratio returns to ≥ 1 (disc build-up/dissipation cycles).

No baseline is yet decisive. Median $\sigma(\mathrm{EW}) \approx 0.44$ Å vs persistence MAE 1.47 Å: about 1 Å is reducible, and PCA(10) + ridge closes only $\sim 8\,\%$ of it.

- Best in-domain: **PCA(10) + ridge, 0.942** ($-5.8\,\%$ MAE).
- **BeMAE** 1.003: EW is near-linear, PCA(10) recovers it exactly.
- Nine zero-shot foundation checkpoints: $\leq 3.2\,\%$ gain; a trained linear head on frozen Chronos-Bolt-B: **0.932**.

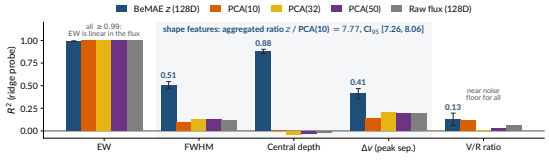
Hα target (R^2)	PCA(10)	PCA(50)	Raw flux
EW	0.66 ± 0.06	0.66 ± 0.06	$\mathbf{0.71 \pm 0.07}$
peak intensity	0.28 ± 0.19	0.38 ± 0.10	0.17 ± 0.56
FWHM	0.00 ± 0.09	0.05 ± 0.02	0.00 ± 0.04
central depth	-0.01 ± 0.07	0.03 ± 0.07	0.00 ± 0.04

Same-night Hβ window $\rightarrow$ Hα features; 2,525 pairs, 1,090 stars (FWHM: 2,508 / 1,086); star-grouped 5-fold CV.

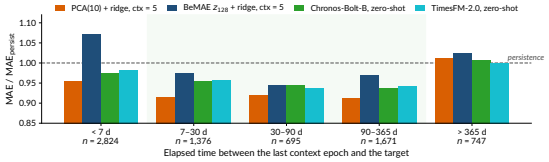
- **Equivalent width transfers**: both lines trace the same disc ($R^2 \approx 0.66$–0.71).
- **Line shape does not**: FWHM and central depth stay at $R^2 \approx 0$, different formation depths.

A measured ceiling for linear single-line readouts.

Line shape: BeMAE wins



EW forecasting: PCA(10) + ridge wins, barely



Take-home message

- **No single winner.** BeMAE for line shape ($\rho = 7.77$), PCA(10) + ridge for EW forecasting (0.942 vs 1.003).
- **Persistence is still the baseline to beat.** Best gain 6.8 %, with a trained head on a frozen foundation model.
- **Open challenges.** Δt-aware forecasting, fine-tuned foundation models, multi-line encoders, symmetry-aware losses for V/R.
- **Plug in yours.** Frozen splits, bootstrap protocol and evaluation pipeline are released.

Paper
OpenReview

Dataset
Hugging Face

Code & pipeline
GitHub

BeMAE weights
Hugging Face

Released under CC-BY-4.0

- ✓ corpus (339,115 spectra) and frozen splits
- ✓ analytic features of the scored Hα slice
- ✓ BeMAE weights, 3 seeds
- ✓ evaluation pipeline and reference results
- ✓ datasheet, Croissant file, Zenodo DOI

Meet us at the poster

Paris Poster Hall · Fri 11 Dec 2026 · 14:30 – 16:30

Contact

matthieu.le-lain@irisa.fr · lelain.net

Snapshot doi:10.5281/zenodo.23808620

Data: BeSS, LIRA, Observatoire de Paris.