Template matching, not time learning: What a self-supervised encoder of star brightness learns

Emma Chickles[1], Kevin Burdge[1], Maggie Sun[1], Pablo Mercader-Perez[1], Jeroen Audenaert[1], Kaley Brauer[2]

[1] MIT [2] Harvard

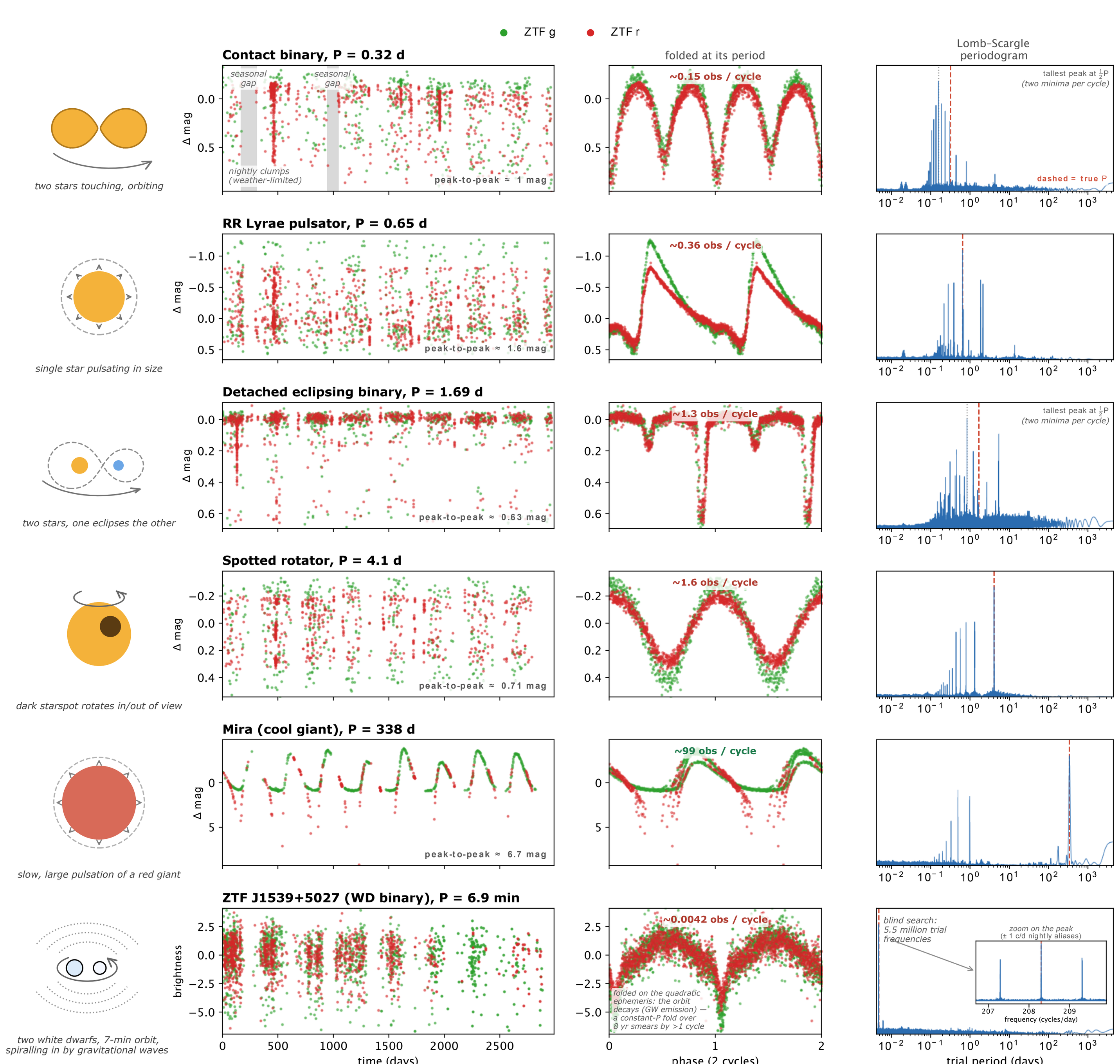
Introduction

Ground-based Zwicky Transient Facility (ZTF; 2018-Present) at Palomar Observatory

- The sky-survey ZTF has monitored ∼2B(illion) stars and galaxies for nearly a decade; Rubin/LSST will extend this to ∼38B (∼17B stars + ∼20B galaxies).
- A variable star's period encodes its physical properties (e.g., temperature, mass, interior structure), calibrates cosmic distances, and traces gravitational-wave-driven orbital decay.
- The Lomb-Scargle periodogram performs a brute-force search over ∼5M trial frequencies per star per decade (2×10^{17} χ^2 evaluations across LSST). **Do self-supervised embeddings measure each star's period — or just learn each class's typical period?**
- A high period R^2 is confounded by class identity. We instead measure a model's ability to read off period from raw time series (light curve):

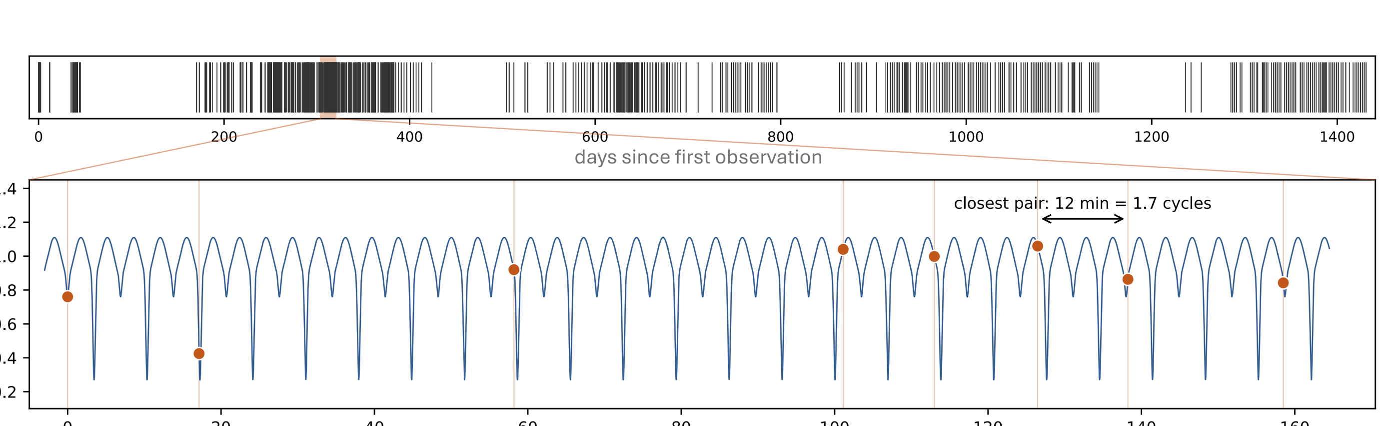
$$\rho_{\text{within}} = \text{median}_c\, \rho_{\text{Spearman}}\,(y, \hat{y}|c), \quad y = \log_{10}(\text{P})$$
$$\text{c} = \text{class}, \quad \text{P} = \text{period}$$

The data: light curves (t_i, f_i, σ^f_i) encode physics



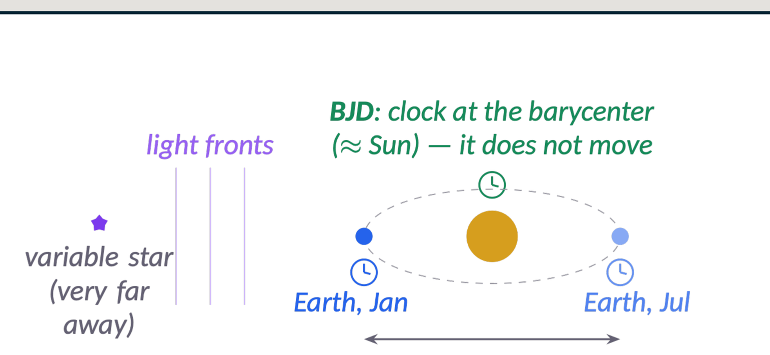
Irregular sampling: beyond the Nyquist Limit

- A typical ZTF field is revisited every ∼2 days in the **g** or **r** filter. For LSST, median revisit of ∼3.5 days across six filters (**u g r i z y**). For example, the 7-minute ZTF J1539+5027 completes ∼410 orbits between ZTF observations, or ∼720 orbits between LSST observations.



- Unlike uniformly sampled signals, irregular observations have no single Nyquist frequency (see VanderPlas 2018), allowing the recovery of short periods like ZTF J1539+5027. However, this isn't for free: recovering it requires searching through a complicated alias structure.

Aside: We use Barycentric Julian Dates (BJD), correcting for Earth's orbital motion, which shifts photon arrival times by up to ±8.3 light-minutes—longer than the 6.9-minute orbit of ZTF J1539+5027. Like numerical precision (float32 vs. float64), the choice of time representation matters.



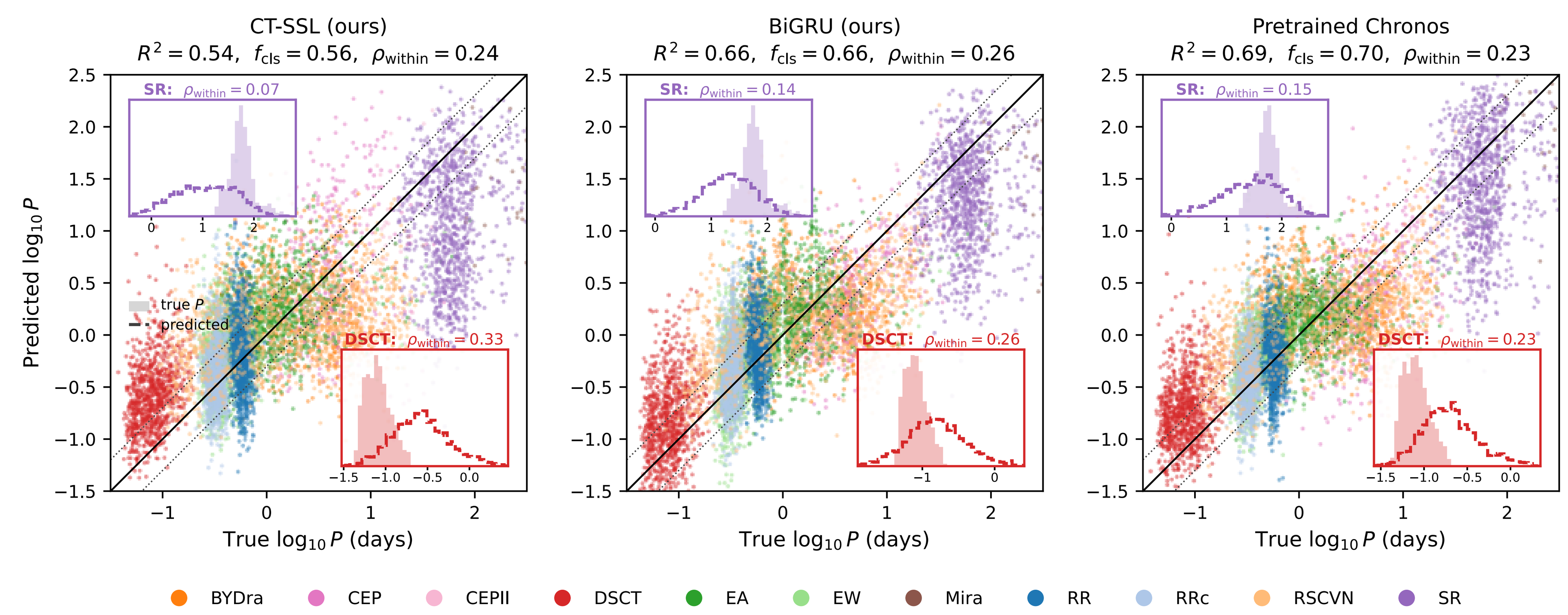
References: Ansari, A.F., et al. 2024, TMLR (Chronos). Audenaert, J., et al. 2025, ICML FMSD Workshop (Casual Foundation Models). Burdge, K.B. et al. 2019, Nature, 571, 528, (ZTF J1539-5027). Donoso-Oliva C., et al. 2023, A&A, 670, A54 (AstroMER). VanderPlas, J.T. 2018, ApJS, 236, 16 (Lomb-Scargle Periodogram).

Do self-supervised representations encode period—or just class?

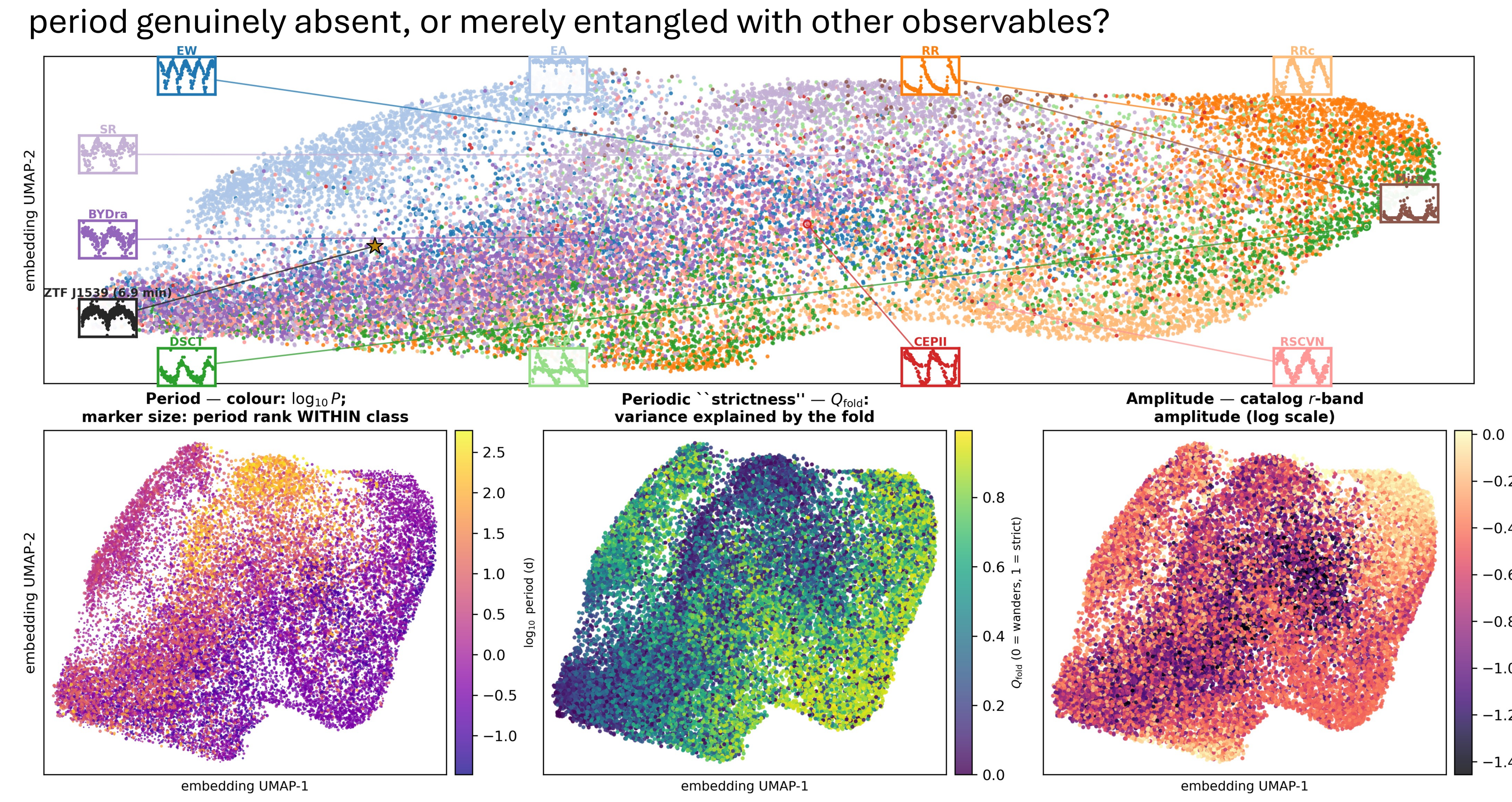
- For each encoder, we freeze the pretrained weights, embed the same 25K labeled light curves, and train an identical linear probe.

		pretraining / input diet	flux	σ	time	how time enters	ρ_{within}
CT-SSL transformer (ours)	4.8M	400k unlabeled ZTF, SSL	✓	✓	✓	Fourier feats + Δt-attention	0.24
BiGRU, cadence-as-channel	4.4M	same 400k ZTF, same SSL	✓	✓	✓	$\log \Delta t$ input channel	0.26
CNN (1D, 6 layers)	4.4M	same 400k ZTF, same SSL	✓	✓	✓	$\log \Delta t$ input channel	0.25
Chronos-T5 (from scratch)	4.9M	400k ZTF, next-token CE	✓	✗	✗	none (token order only)	0.04
Chronos-T5-small (frozen)	46M	$\sim 10^9$ generic timesteps	✓	✗	✗	none	0.23
Chronos-T5-small + ZTF	46M	generic $\to$ 120k ZTF fine-tune	✓	✗	✗	none	0.22
MOMENT-base (frozen)	110M	$\sim 10^9$ generic timesteps	✓	✗	✗	assumes a regular grid	0.21
ASTROMER (frozen)	0.7M	1.5M MACHO R-band LCs, masked SSL	✓	✗	✓	time-based positional encoding	0.12
Lomb-Scargle + argmax	—	—	✓	✗	✓	explicit: ∼5M trial folds	**0.78**

- Classes clearly separate in embedding space, but period does not; every encoder achieves $\rho_{\text{within}} \leq 0.26$ (128-1024-d embeddings — the standardized ridge probe cross-validates its regularization per encoder, and an MLP probe gives the same ordering). In contrast, a naïve argmax on Lomb-Scargle periodograms (the classical tool) reaches $\rho_{\text{within}} = 0.78$.



- Coloring the embedding reveals smooths trends in periodic strictness (variance of phase-folded profile) and variability amplitude, both within and across variability classes.
- These properties covary with period (e.g., RR Lyrae amplitude anti-correlates with period): is period genuinely absent, or merely entangled with other observables?



Period search computation is hard to learn for irregular sampling

- We observe that a neural network's ability to read off a period from a given a light curve depends on sampling cadence relative to the intrinsic periodicity (*obs/cycle*).
- **Left:** Under dense sampling (10-100 *obs/cycle*), a network predicts nearly perfectly; under sparse sampling (0.2-2 *obs/cycle*), period information disappears. *(Synthetic signals, real ZTF timestamps.)*
- **Middle:** When easy examples are present, period learning nucleates at long periods and propagates toward shorter periods during training. With sparse-only data, learning never begins.
- **Right:** Loss-landscape measurements show a qualitative transition: when period learning emerges, the optimization floor collapses (∼10×) and curvature increase.