THE UNIVERSITY OF CHICAGO DATA SCIENCE INSTITUTE

The University of Chicago AI+Science Summer School

July 17-21, 2023

Attention-based Modeling of AGN Variability

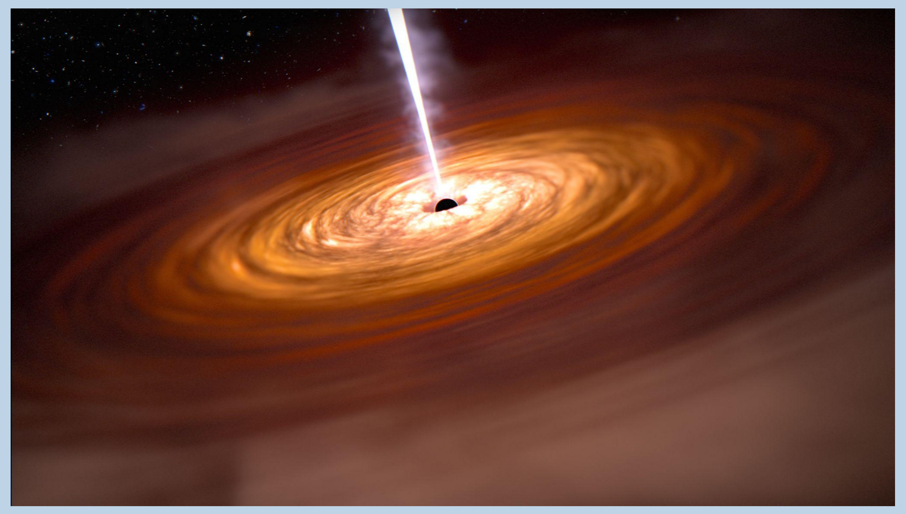
Matthew Lowery, Advised by Dr. Hermine Landt-Wilman

Durham University

Durham Centre for Extragalactic Astronomy

introduction

- Active Galactic Nuclei (AGN) a.k.a. accreting supermassive black holes are some of the most luminous objects in the universe and the energy they radiate greatly affects the galaxies that host them.
- Understanding the accreting region's underlying physics and resultant geometries then would be insightful, but such regions are generally too compact, too far away and/or too obfuscated by gas/dust to be spatially resolved.



Simulated depiction of AGN, retrieved from: https://www.nasa.gov/feature/goddard/2021/simulated-webb-images-of-quasar-and-galaxy-surrounding-quasar

dataset

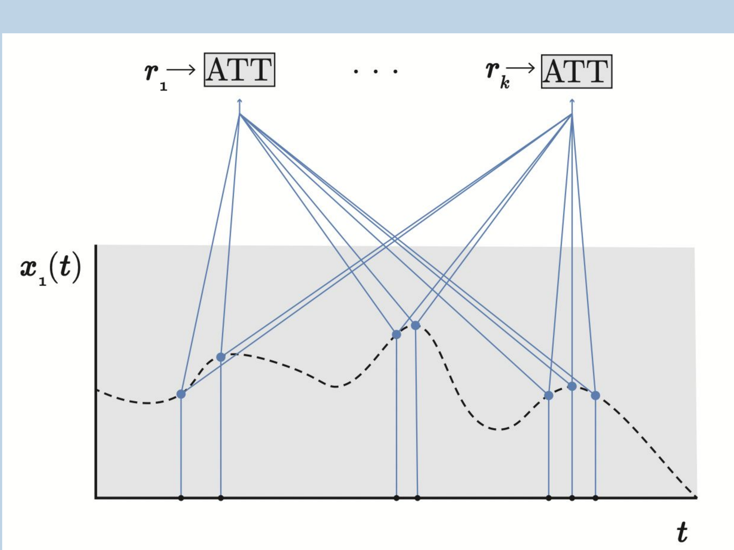
- The most ubiquitous and easy-collected data for AGN are **photometric light curves**, where the data points reflect all light collected between a given wavelength range (corresponding to the filter used) at any given time, taking the form of **irregularly observed time series**.
- We utilize light curves from the Zwicky Transient Facility (ZTF; [1]), which is located at the Palomar Observatory and scans the entire Northern sky with an approximately three night cadence for for its custom *g* and *r* filters and a less resolved four night cadence for its *i* filter.
- ZTF's api was queried based on a sample catalog with 3398 objects generously provided by Paula Sánchez Sáez to obtain light curve files for each of the bands from data release 15, which had a total observation span from March 2018 – November 2022.

model: HeTVAE [2]

- Light curves are a **projection of the activity and observed variability patterns of the accreting black holes** so intuiting the data generating process behind them (via a model) gives way to intuiting more about the underlying processes they are resultant of.
- The difficulty in modeling this data is both in its **stochasticity and its irregularity**, which means non-trivial accommodations need be made for neural networks to ingest the data.
- The most prominently used model for these light curves is a Damped Random Walk (DRW; [3]), which makes assumptions of AGN variability patterns and inappropriately constrains light curve information to 2 parameters (using Gaussian processes).
- As such, we consider Heteroscedastic Temporal Variational Autoencoder (HeTVAE; [2]), which is an attention-based neural network that makes no such assumptions and was originally test on medical data, having the functionality of creating a summary distribution of our light curves and probabilistically interpolating an irregularly sequenced, multivariate time series by querying this summary distribution/latent representation.

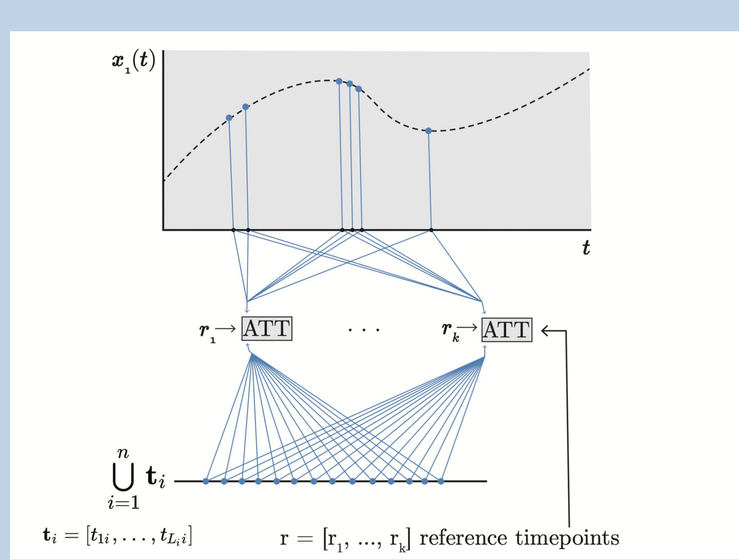
attention modules

- To discretize the irregular time series the network attempts **to project information to a series of reference time points** through attention modules, coined ***value*** and ***intensity***.
- Attention, computed via a familiar scaled dot product compatibility function, is taken between reference time points and observed time points after such values are are projected to a higher dimensional plane using a dense layer with a sinusoidal activation.
- Each projection is relative to a particular attention head, akin to multi-head attention [4].



Depiction of the ***Value*** encoding. Attention is taken between a reference time point and an example's observed time points and softmaxed. These scores are weighted by the observed values x(t) and summed to create a representation x for the reference time point.

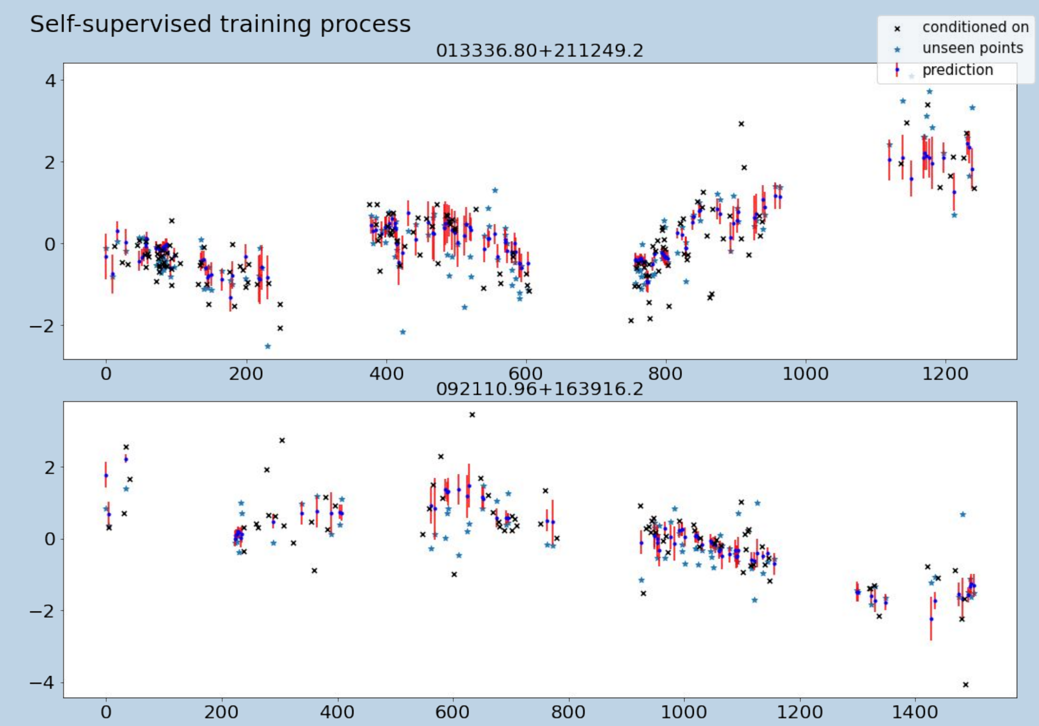
Hint: The same calculations to project the light curve's information to the reference time points can be done to interpolate/forecast once the network is trained, where instead we project this information to our choice of time points from the information that is encapsulated in the reference time points.

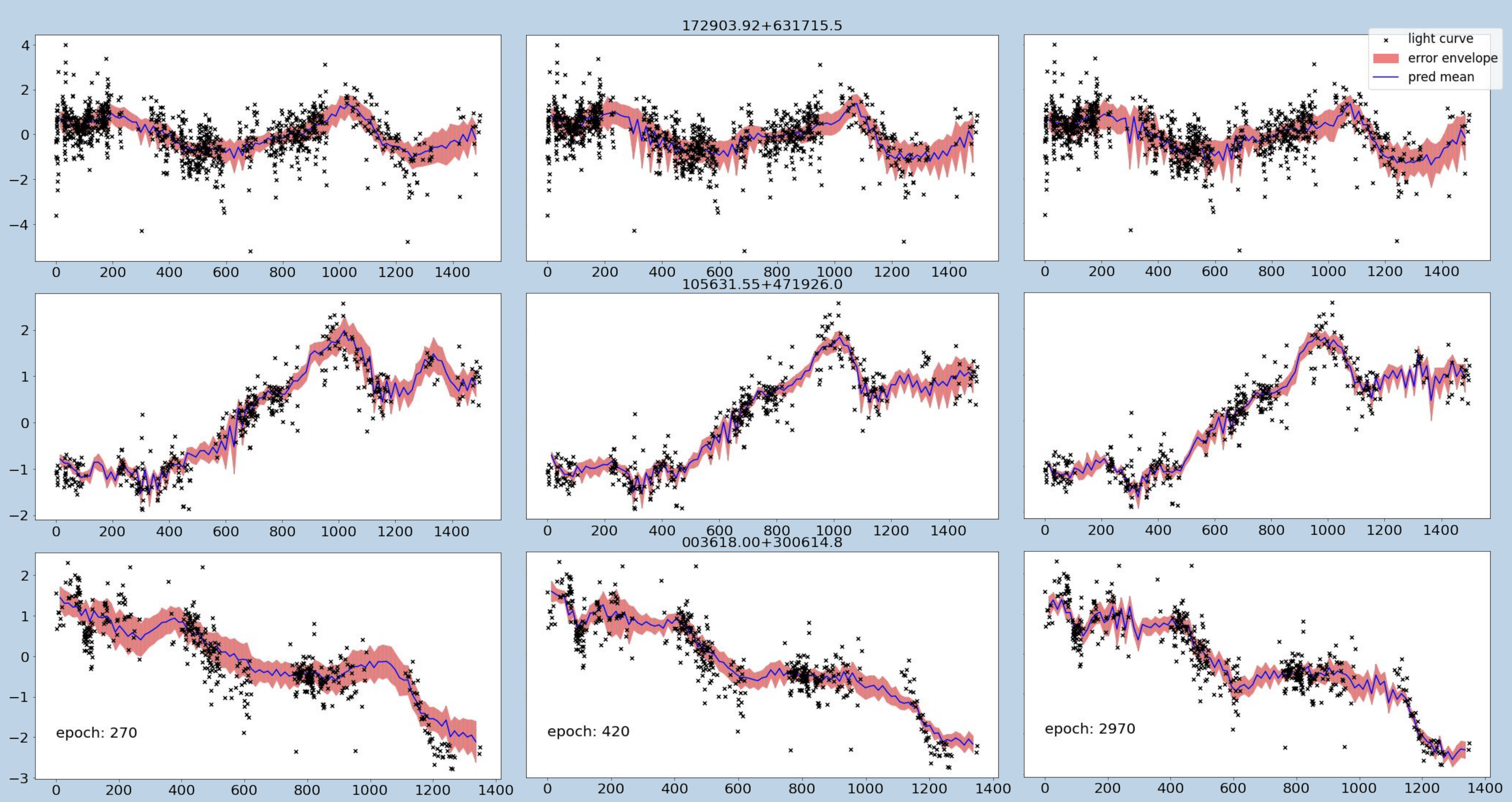


The ***Intensity*** encoding is a way to project observational sparsity information. Attention is calculated between a reference time point and the observed time points as well as between the same reference time point and the union of all time points in the dataset and after pooling each set of calculations, the ratio of the two is taken. Intuitively, if attention is large between a reference time point and an observation relative to the union of all observed time points across the dataset, there is less observational sparsity at that reference point.

train it!

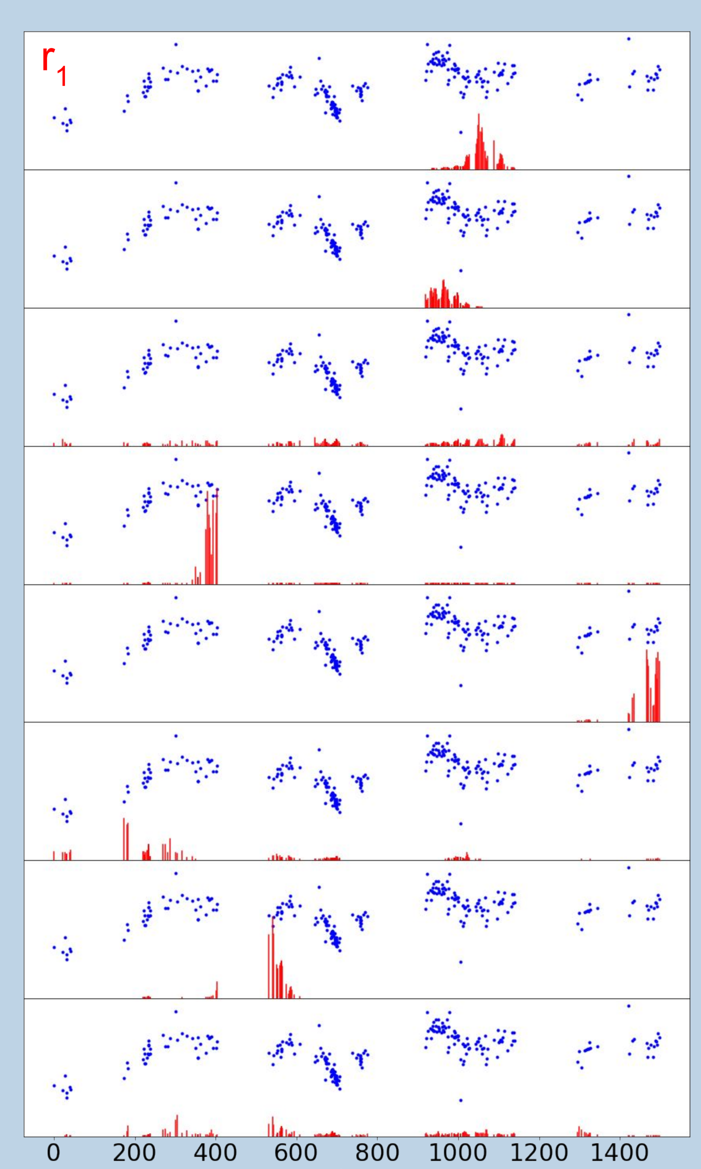
- HeTVAE employs a **self-supervised training regimen** to read semantic information from the light curves.
- This procedure is under the umbrella of unsupervised learning in so much as it uses the structure of the data itself to learn, but its technique works by hiding a subsample of the data from the network and making predictions for specifically that hidden data (shown on the right).
- We apply a cyclical KL annealing schedule while optimizing the VAE's ELBO objective function to mitigate KL vanishing.

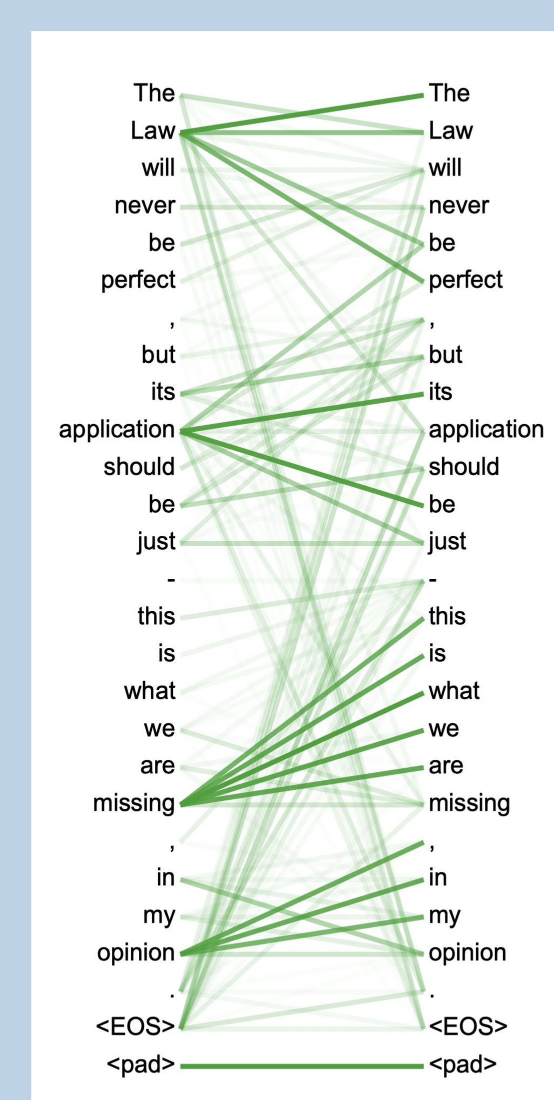




Example interpolations for three arbitrary objects as training progresses to 100 points evenly spread between each light curve's range. The title shows the object identifier, i.e. its location in the sky the Julian equatorial coordinate system. On the y-axis are normalized magnitude values; on the x-axis is days starting from 0.

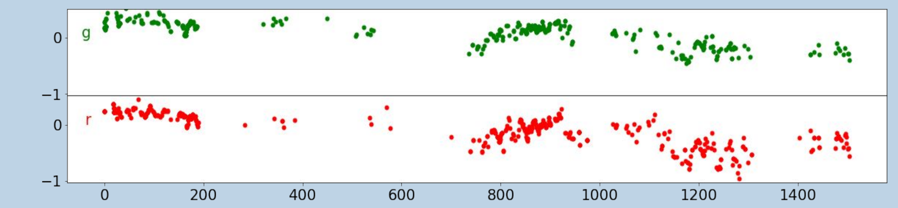
visualize attention scores from ***value*** encoding





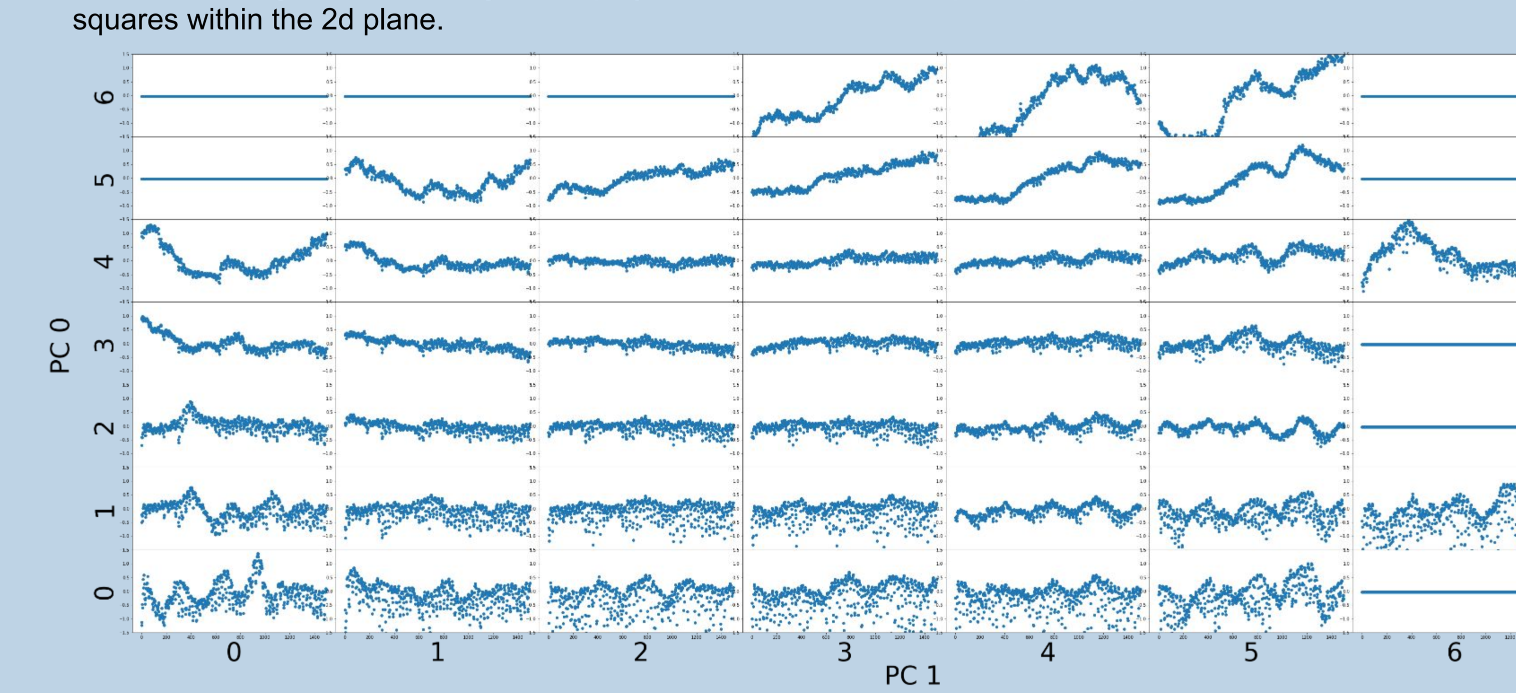
- On the left we show normalized attention scores from the first reference point for the *value* encoding across 8 attention heads (top to bottom) using an arbitrary *g* filter light curve from the dataset. **How much attention the reference point pays to a given observation is reflected by by the length of the red lines.**
- On the right we show an analogous visual of the attention scores from *Attention Is All You Need* [4], where instead scores (from self-attention) are reflected by the opacity of the lines.

Bonus! Create synthetic data by sampling the latent space. Here we project the reconstruction onto time points from an example in the our dataset.



interpret how light curves are distributed across the latent space

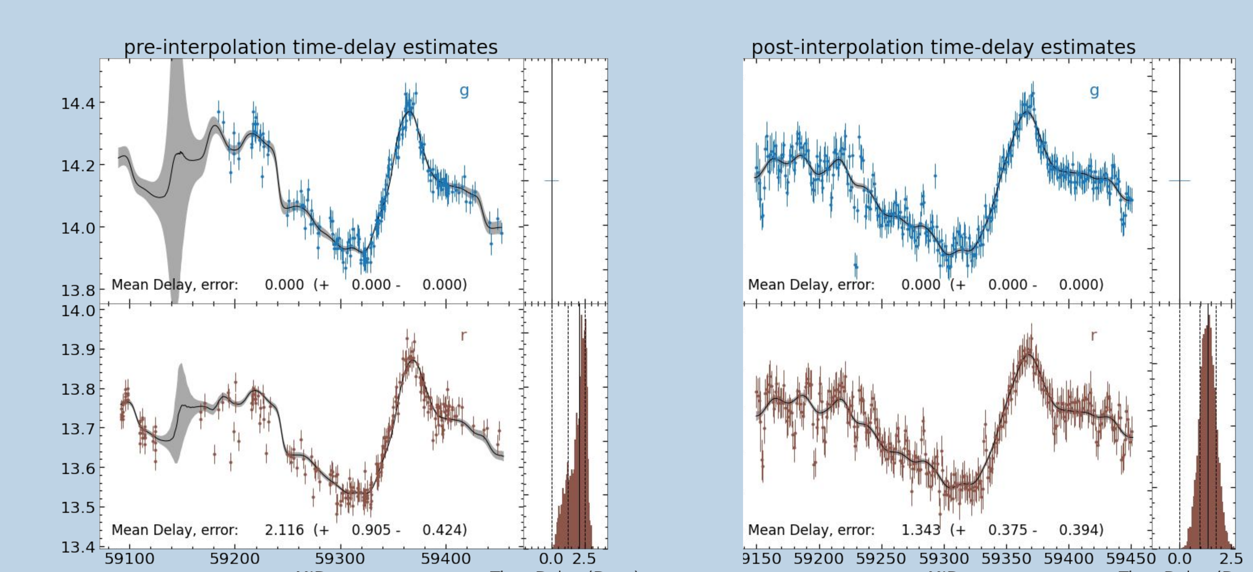
- In this experiment, we use PCA to help us discern how light curves with different variability patterns are distributed in the latent space. To see what variances in the latent space the principal components are pointing out, which should point out intuited `differences' in variability of the light curves, **we can bin the light curve's embeddings after they are projected onto some of these components, average the embeddings within the bins and reconstruct those averages to see what the reconstruction looks like**.
- To create this figure, embeddings were projected onto the first two principal components so that the bins were squares within the 2d plane.



- The middle bins reflect the mass population of light curves, which reveals that the network is very biased/the data is very imbalanced with respect to light curves that show minimal variability fluctuations. Because more variable light curves are more interesting to study and we don't want performance to be worse for them/ for them to all appear anomalous, we would likely mitigate the imbalance in the future (using the VAE itself or using variance metrics) before fitting a model, although we still could use this one as a means to filter out variable light curves from large datasets.
- **Note**: If we were looking for more truthfully anomalous AGN, we likely would not want returned anomalies with respect to any parameter we can already calculate (that affects observed variability) so we would want to correct for any such biases/dataset imbalances first.

apply interpolations to reverberation mapping

- In *photometric reverberation mapping*, we aim to **better constrain AGN geometries by calculating a time-delay *(τ)* between light curves of different photometric filters**, which is proportional to the distance between the regions in the AGN from which the light of the different light curves emanates (because the speed of light is constant). As such, the radius ~ c*τ*.
- In practice, time-delays are estimated by a cross-correlation technique between two light curves, i.e. scoot the light curve that corresponds to a higher energy (and thus likely closer to the black hole) photometric filter over in time until the correlation between the two light curves is maximized, i.e. when their shapes best match.
- The difficulty is that light curves are irregularly observed, which makes such an analysis difficult. In general, current efforts use bayesian running optimal averages [5] or the aforementioned DRW [3] to interpolate the light curves before cross-correlating them.
- Thus, we instead can use multivariate interpolations from HeTVAE to improve time-delay estimates (we fine-tune the interpolations for the light curves associated with specific objects).



<u>Aiming for:</u>

0.93 (+0.06 -0.07) days.

- In this case, we have access to a known time-delay for the same object (the AGN NGC 5548) computed from much higher cadence light curves than our ZTF data.
- *g* is the higher energy filter so we calculate the *r* filter light curve's time-delay.
- We use a modern algorithm called PyROA [5] (hint it's a ROA) to try to estimate the delays pre-interpolation (left; est. 2.116 (+0.905 -0.424) days) and post-interpolation (right; est. 1.343 (+0.375 -0.394) days) with our ZTF light curves and thus were able to estimate closer to the known reference time-delay, which is: **0.93 (+0.06 -0.07) days.**

References

- [1] Masci, F. J., Laher, R. R., Rusholme, B., et al. 2018, *The Zwicky Transient Facility: Data Processing, Products, and Archive*, PASP, 131, 995.
- [2] Satya Narayan Shukla and Benjamin M. Marlin. Heteroscedastic Temporal Variational Autoencoder For Irregularly Sampled Time Series. 2021. doi: 10.48550/ARXIV.2107.11350. url: https://arxiv.org/abs/2107.11350.
- [3] B. C. Kelly, J. Bechtold, and A. Siemiginowska, "Are the Variations in Quasar Optical Flux Driven by Thermal Fluctuations?," *The Astrophysical Journal*, vol. 698, pp. 895–910, June 2009.
- [4] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," CoRR, vol. abs/1706.03762, 2017.
- [5] F. R. Donnan, K. Horne, and J. V. Hernández Santisteban, "Bayesian analysis of quasar light curves with a running optimal average: new time delay measurements of COSMOGRAIL gravitationally lensed quasars," Monthly Notices of the Royal Astronomical Society, vol. 508, pp. 5449–5467, Dec. 2021.