



Reducing the Dimensions of Active Galactic Nuclei Light-curve Manifolds

Shoubaneh Hemmati¹ , Jessica Krick¹ , Daniel Stern² , Vandana Desai¹ , Andreas Faisst¹ , Lucas Martín-García³ , Varoujan Gorjian¹ , Aryana Haghjoo⁴ , Farnik Nikakhtar⁵ , Troy Raen¹ , Sogol Sanjaripour⁴ , Brigitta M. Sipőcz¹ , and

David Shupe¹

¹IPAC, California Institute of Technology, Pasadena, CA 91125, USA; shemmati@caltech.edu

²Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA

³Universidad Rey Juan Carlos, Mostoles, Madrid 28933, Spain

⁴Department of Physics and Astronomy, University of California Riverside, Riverside, CA 92521, USA

⁵Yale University, New Haven, CT 06511, USA

Received 2025 May 12; revised 2026 January 10; accepted 2026 January 12; published 2026 February 5

Abstract

The active galactic nuclei (AGNs) glossary is vast and complex. Depending on the selection method, observing wavelength, and brightness, AGNs are assigned distinct labels, yet the relationship between different selection methods and the diversity of time-domain behavior within and across classes remains difficult to characterize in a unified framework. Changing-look AGNs (CLAGNs), which transition between classifications over time, further complicate this picture. In this work, we learn a data-driven, low-dimensional representation of multiwavelength photometric light curves of AGNs, in which the structure of the projected manifold correlates with AGN class and independent spectroscopic properties. Using the NASA Fornax Science Platform, we assemble light curves from Zwicky Transient Facility, Pan-STARRS, Gaia, and Wide-field Infrared Survey Explorer/NEOWISE for two samples: (1) a heterogeneous set of ~ 2000 AGNs spanning $z \lesssim 1$, including Sloan Digital Sky Survey quasars, variability-selected sources, and CLAGNs; and (2) a homogeneous sample of $\sim 65,000$ narrow-line AGNs at $z \approx 0.1$ with well-characterized optical emission-line measurements. Without using class labels during training, the learned manifolds organize variability-selected AGNs into coherent regions of the low-dimensional space, distinguish between turn-on and turn-off CLAGNs, and place tidal disruption events in distinct regions. Manifold coordinates correlate with key spectroscopic and host-galaxy properties—including stellar mass, [O III] luminosity, and $D_n(4000)$ —demonstrating that heterogeneous multiband variability can be combined in a purely data-driven manner to recover correlations with independent physical diagnostics, without requiring explicit physical modeling. These results show that manifold learning offers a practical, assumption-light approach for integrating time-domain surveys and prioritizing spectroscopic follow-up.

Unified Astronomy Thesaurus concepts: [Active galactic nuclei \(16\)](#)

1. Introduction

Active galactic nuclei (AGNs) are among the most luminous and variable sources in the Universe, powered by accretion onto supermassive black holes. In the unified model, the diversity of AGN observational classes is primarily explained by orientation effects and obscuration rather than intrinsic physical differences (e.g., R. Antonucci 1993; C. M. Urry & P. Padovani 1995; H. Netzer 2015). For example, type-1 AGNs, with a direct view of the central engine, exhibit broad emission lines in their optical spectra, whereas type-2 AGNs have these regions obscured by a dusty torus, showing only narrow emission lines. This orientation-based framework has been successful in explaining a wide range of observed features, including the presence of polarized broad lines in some type-2 sources (R. R. J. Antonucci 1984). However, it treats AGN classification as static and does not incorporate temporal variability or evolutionary changes within individual sources.

A defining property of AGNs is their variability across all wavelengths (M.-H. Ulrich et al. 1997). Optical and UV continuum fluxes vary stochastically by tens of percent over

timescales from days to years, and X-ray fluxes can change even more rapidly (e.g., I. M. McHardy et al. 2006; S. Koshida et al. 2014). These fluctuations reflect dynamic processes in the accretion flow and the immediate surroundings of the black hole (e.g., B. C. Kelly et al. 2009; C. L. MacLeod et al. 2010; C. J. Burke et al. 2021). Reverberation mapping exploits this variability to estimate the size of the broad-line region and infer black hole masses (e.g., B. M. Peterson 2001; M. C. Bentz et al. 2013). In the optical, host galaxy starlight can dilute variability, whereas UV bands trace thermal disk fluctuations, and IR variability reflects reprocessed emission from the dusty torus (B. J. Shappee et al. 2014; E. M. Cackett et al. 2021). These differences highlight the multiscale nature of AGN variability and motivate multiband studies of time-domain behavior.

One class of AGNs that challenges static classification schemes is the changing-look AGNs (CLAGNs), which switch between type-1 and type-2 spectral states or undergo “turn-on” and “turn-off” episodes on observable timescales. These transitions, marked by the appearance or disappearance of broad emission lines and significant continuum variability, are generally attributed to changes in accretion rate or obscuration along the line of sight (e.g., S. M. LaMassa et al. 2015; J. J. Ruan et al. 2016; D. Stern et al. 2018; Q. Yang et al. 2018; M. J. Graham et al. 2020; C. Ricci & B. Trakhtenbrot 2023). The first spectroscopically confirmed CLAGN was reported by



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

S. M. LaMassa et al. (2015), revealing a decline in optical and X-ray flux accompanied by the loss of broad $H\alpha$ emission. Since then, time-domain spectroscopic surveys have significantly expanded the known population, with multiepoch observations confirming changes in Balmer line profiles and continuum luminosity across a growing number of sources (e.g., M. J. Graham et al. 2020; P. J. Green et al. 2022; Q. Yang et al. 2023, 2025; S. Wang et al. 2025). Mid-infrared and optical diagnostics have been used to disentangle intrinsic changes in accretion from variable obscuration, with current evidence often favoring accretion-driven mechanisms (e.g., Q. Yang et al. 2018; N. P. Ross et al. 2020).

Machine learning (ML) techniques have recently been adopted for CLAGN discovery and characterization. Supervised models, such as random forest classifiers, have been trained on optical variability features to identify numerous candidates, many subsequently confirmed through spectroscopy (e.g., E. López-Navas et al. 2022). Deep learning approaches have also been applied to capture complex light-curve structure and distinguish CLAGNs from other variable AGN or transient populations (e.g., P. Sánchez-Sáez et al. 2021). More broadly, ML offers scalable approaches for organizing AGN variability and identifying candidate classes, though most existing efforts rely on individual surveys rather than integrating multiwavelength time-domain data.

In this work, we use dimensionality reduction on AGN time-series data drawn from multiple photometric archives, accessed via the NASA Fornax Science Platform. We process heterogeneous light curves into a consistent multiband representation and construct a high-dimensional description of time-domain photometric behavior. We then apply non-linear dimensionality reduction to organize sources by similarity in their multiband variability. Manifold learning has proven effective in studies of stellar variability (e.g., K. V. Sokolovsky et al. 2017), AGN light curves (e.g., A. L. Faisst et al. 2019; R. Pantoja et al. 2022), and galaxy spectral energy distributions (e.g., D. Masters et al. 2015; S. Hemmati et al. 2019a, 2019b; S. Sanjaripour et al. 2024, 2025). By projecting AGN subpopulations into a reduced-dimensional space, groupings, transitions, and outliers in time-domain variability become more readily apparent, providing an interpretable framework for comparison across samples.

Working with asynchronous, multiband light curves remains a significant challenge in time-domain astronomy. Several recent efforts have developed statistically rigorous models for this task, including multivariate Gaussian processes (GPs) for modeling light curves across bands (C. Gilbertson et al. 2020), autoregressive models adapted for irregular time series (Z. Hu & H. Tak 2020), and GP factor analysis for recovering latent variability structure (F. Elorrieta et al. 2021). These methods aim to model the underlying generative processes and perform interpolation under explicit statistical assumptions. In contrast, our approach uses observed light-curve behavior directly as input for dimensionality reduction, without attempting to model the generative variability process. Rather than focusing on prediction or imputation of missing data, our goal is to construct a similarity-based representation that captures dominant patterns in multiband time-domain variability across diverse AGN populations.

We test this approach on two distinct AGN samples. The first is a heterogeneous compilation of AGNs spanning up to

$z \sim 1$, including Sloan Digital Sky Survey (SDSS) quasars, Wide-field Infrared Survey Explorer (WISE)- and Galaxy Evolution Explorer (GALEX)-selected variable sources, and CLAGNs drawn from the literature. This sample is used to assess how labeled, variability-selected populations distribute within the learned time-domain space when labels are not used during training, testing whether population distinctions identified in individual surveys persist when all multiband photometric light curves are analyzed jointly. The second sample is a homogeneous set of type-2 AGNs at $z \sim 0.1$ with well-characterized optical emission-line diagnostics (G. Kauffmann et al. 2003). This sample provides extensive spectroscopic measurements, including [O III] luminosity, $H\delta$ absorption, and $D_n(4000)$, enabling a direct comparison between time-domain variability structure and independent host-galaxy and nuclear properties. Its uniform redshift coverage and clean classification make it a valuable reference set, validating whether manifold coordinates learned from photometric variability alone exhibit systematic trends with independently measured spectroscopic and host-galaxy properties.

The remainder of this paper is organized as follows. In Section 2, we describe the time-series data and the Fornax platform. Preprocessing is presented in Section 3, and the two AGN samples are introduced in Section 4. The manifold learning methodology is described in Section 5, and results are presented in Section 6. We discuss limitations and implications and summarize our conclusions in Section 7.

2. Archival Time-series Data

As part of the NASA Fornax Initiative, we have developed and utilized the Fornax Science Platform to query and process time-series data from multiple NASA and non-NASA archives. Fornax is a cloud-based science platform designed to integrate data access, software environments, and scalable computing resources, enabling reproducible analysis of large, heterogeneous astronomical datasets. Throughout this work, Fornax is used as an access and processing framework rather than as a scientific method in itself. The Jupyter notebooks used to retrieve and preprocess the light curves analyzed in this paper are publicly available.⁶

This study incorporates time-domain photometric data from the following surveys: the Zwicky Transient Facility (ZTF; E. C. Bellm et al. 2019), WISE/NEOWISE (E. L. Wright et al. 2010; A. Mainzer et al. 2011), Panoramic Survey Telescope and Rapid Response System (Pan-STARRS; K. C. Chambers et al. 2016), and Gaia (Gaia Collaboration et al. 2016). While additional time-domain datasets—such as those available through HEASARC (e.g., Fermi and BeppoSAX) or IceCube—could, in principle, be integrated into the same framework, we do not include them here due to differences in cadence, temporal baseline, wavelength regime, or sky coverage that would complicate direct comparison within the scope of this work.

ZTF is a wide-field, high-cadence optical survey conducted at the Palomar Observatory, covering approximately 25,000–30,000 square degrees of the northern sky in three filters (g , r , and i). For this analysis, we use the processed and cataloged DR18 photometry, accessed via positional matching

⁶ See *Fornax Lightcurve Notebooks* at <https://nasa-fornax.github.io/fornax-demo-notebooks/ml-agnzoo/>.

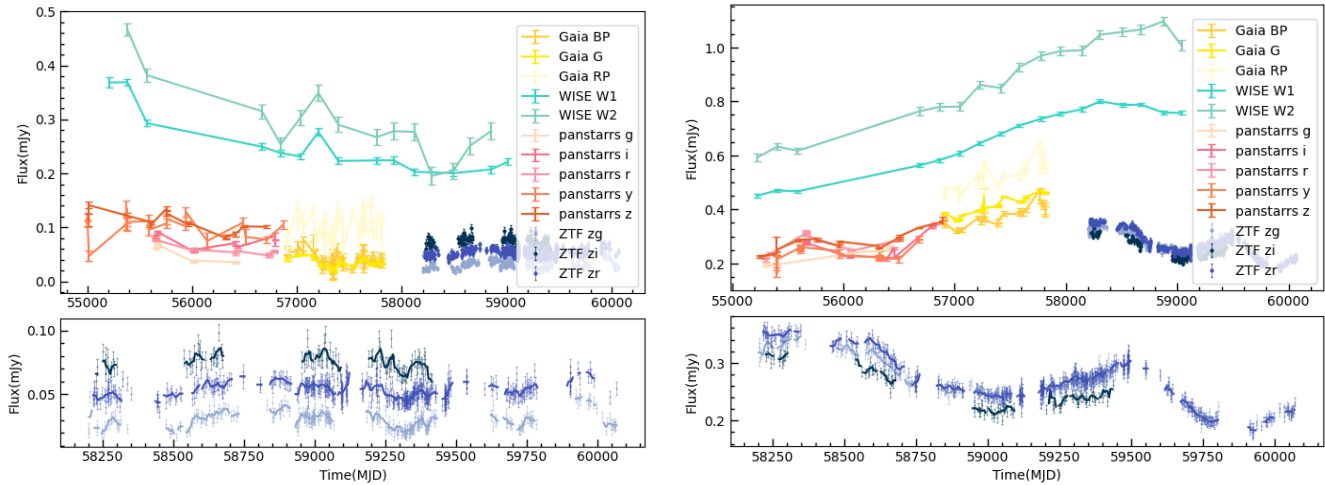


Figure 1. Light curves of two AGNs at $z = 0.277$ and $z = 0.79$, queried from Gaia, ZTF, Pan-STARRS, and WISE, shown in the left and right panels, respectively. The bottom panels are zoomed-in views of the ZTF bands.

with a $1''$ radius, rather than starting from image-level data. The ZTF data reduction and calibration pipeline is described in detail by F. J. Masci et al. (2019).

Pan-STARRS operates a 1.8 m telescope in Hawaii and observes the sky north of decl. $\sim -30^\circ$ in five filters (g , r , i , z , and y), with typically ~ 12 epochs per filter. We retrieve Pan-STARRS photometry through cone searches with a $1''$ radius on the publicly released catalogs hosted on MAST. The survey design, data processing, and calibration procedures are described by H. A. Flewelling et al. (2020).

Gaia is an ESA space mission that provides optical time-series photometry in three bands: a broad G band and two narrower bands, G_{BP} and G_{RP} . Gaia DR3 includes time-series data for approximately 11 million variable and nonvariable sources. We retrieve Gaia light curves from the Gaia DR3 *source_lite* catalog (Gaia Collaboration et al. 2023) using *astroquery*. The *source_lite* table provides a streamlined subset of the full *gaia_source* catalog optimized for large-scale cross-matching and light-curve access. We join uploaded target lists with the Gaia photometry tables to obtain multiepoch flux measurements.

In the mid-infrared, WISE and NEOWISE provide all-sky coverage with a temporal baseline exceeding a decade and a typical cadence of ~ 12 visits per sky location per epoch. We use the W1 ($3.4 \mu\text{m}$) and W2 ($4.6 \mu\text{m}$) bands from the WISE unTimely catalog (A. M. Meisner et al. 2023). The WISE photometry is accessed from publicly available Amazon Web Services Open Data Repository buckets, stored in Parquet format and spatially partitioned using HEALPix indexing (K. M. Górski et al. 2005), enabling efficient large-scale queries.

For each object, the queried photometry is stored in a Pandas multi-index DataFrame containing the object identifier, redshift (when available), an external label describing the object’s selection class (e.g., WISE-variable AGN or CLAGN), and the photometric measurements and uncertainties in each band. All fluxes are converted to consistent units (μJy). The resulting dataset consists of asynchronous, irregularly sampled multiband light curves with heterogeneous cadence and temporal coverage across surveys. Figure 1 shows the light curves of two example quasars at $z = 0.27$ and $z = 0.79$.

3. Preprocessing of Light Curves

Although the light-curve data are stored in a consistent format and unit, additional preprocessing is required to enable a combined analysis across surveys and wavelength bands. Steps such as aligning time grids, handling missing data, choosing an effective temporal resolution, and applying flux normalization inevitably involve methodological choices that shape how variability is represented in the final analysis.

Because the light curves considered here are asynchronous and heterogeneous in cadence, depth, and temporal coverage across bands and surveys, many standard ML methods cannot be applied directly. While several statistically rigorous approaches exist for modeling or imputing irregular time series (e.g., C. Gilbertson et al. 2020; Z. Hu & H. Tak 2020; F. Elorrieta et al. 2021), our goal at this stage is not to model the underlying generative variability process or to optimize interpolation performance. Instead, we adopt simple and transparent preprocessing strategies that are sufficient to place heterogeneous light curves on a common footing for similarity-based, distance-driven manifold learning. A comprehensive comparison to alternative imputation or time-series modeling frameworks is therefore beyond the scope of this work. For context, univariate time-series imputation methods have been extensively studied in the statistical literature (e.g., S. Moritz et al. 2015; W. Beck et al. 2018).

Most ML algorithms operate on fixed-length inputs and cannot be applied directly to irregularly sampled light curves. Because our dataset combines multiple bands from multiple observatories, each with its own cadence, we unify the time grids by interpolating each band onto a common temporal grid. We explore two interpolation strategies: nearest-neighbor linear interpolation and GP regression (C. Williams & C. Rasmussen 1995) with a rational quadratic kernel. This kernel is chosen for its ability to represent variability across a range of characteristic timescales, effectively acting as a scale mixture of squared-exponential kernels. Figure 2 illustrates both interpolation methods for representative ZTF and WISE bands.

Nearest-neighbor interpolation provides a simple and transparent baseline that preserves observed values without introducing additional model assumptions. GP regression, in

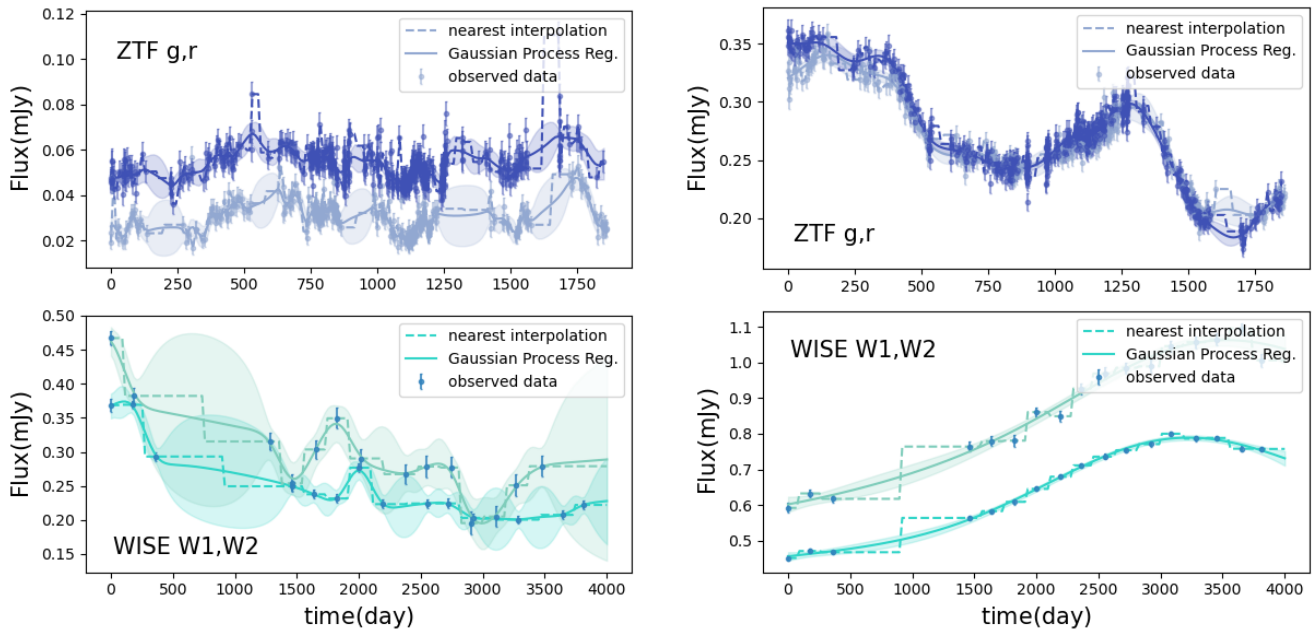


Figure 2. Unified time arrays for the same two objects shown in Figure 1. The ZTF g and r bands, along with the WISE W1 and W2 bands, are displayed. Gaussian process regression is shown with solid lines and shaded regions, and nearest-neighbor linear interpolation is depicted with dashed lines.

contrast, offers greater flexibility in modeling correlated variability and interpolating across gaps, while naturally incorporating measurement uncertainties through the covariance structure. GP regression is therefore well-suited for heterogeneous astronomical time series, albeit at substantially higher computational cost. We retain both approaches in our initial analysis to illustrate sensitivity to interpolation strategy and to motivate our choice of GP-based unification for the remainder of the work.

Feature scaling is a standard preprocessing step for many ML algorithms, particularly those based on distance metrics or gradient-based optimization. In the context of AGN light curves, the choice of normalization is nontrivial, as it should suppress trivial luminosity differences while preserving physically relevant variability patterns and relative flux information across bands. To this end, we normalize each object’s multiband light curve by dividing all bands by the 1σ clipped maximum flux of the band with the largest number of observations. This choice ensures that the reference band is well sampled, reduces sensitivity to outliers, and preserves relative flux ratios between bands, thereby maintaining color information in the normalized light curves. We have verified that the qualitative structure of the learned manifold is robust to reasonable variations in normalization strategy, such as using a clipped median instead of a clipped maximum or selecting a different reference band, provided that relative flux structure across bands is preserved.

Finally, we apply a simple temporal alignment by shifting the start time of each light curve to zero (Figure 2). This standardization allows light curves of different lengths and absolute epochs to be compared within a unified framework and is commonly used when the analysis focuses on variability patterns rather than absolute timing (e.g., D. Folgado et al. 2018). We emphasize that this choice intentionally discards absolute timing information and is therefore not designed to capture inter-band time lags or reverberation effects. Investigating relative timing and lag structures in a multiband

framework is an important direction for future work, but lies beyond the scope of the present analysis.

4. Two Distinct Samples from the Literature

To explore the structure of the multiband AGN light curves in a data-driven manner, we compile two deliberately distinct samples from the literature, each designed to probe complementary aspects of AGN variability and population diversity. Sample A is a heterogeneous compilation extending to $z \sim 1$, assembled to test whether unsupervised analysis of multiband light curves can recover known variability-selected populations, including CLAGNs and tidal disruption events (TDEs). Sample B, in contrast, is a homogeneous, low-redshift ($z \sim 0.1$) sample of narrow-line (type-2) AGNs with uniformly measured spectroscopic and host-galaxy diagnostics, enabling a direct comparison between variability and independent physical observables. These two samples serve different purposes in the analysis. Sample A is used to assess how labeled variability-selected populations project into the learned manifold when labels are not used during training. Sample B is used to validate whether the manifold learned from photometric time series alone exhibits systematic trends with well-established spectroscopic and host-galaxy properties.

4.1. Variables and CLAGNs (Sample A)

The left panel of Figure 3 summarizes the components of sample A and their redshift distributions. This sample consists of AGNs drawn from multiple surveys and selection strategies, and individual objects may appear under more than one label. For example, an AGN may be both X-ray selected and identified as variable in WISE. Labels are therefore not mutually exclusive and are used solely for post hoc interpretation of the manifold.

Approximately half of sample A is drawn from the SDSS DR16 catalog of spectroscopically confirmed quasars (B. W. Lyke et al. 2020), labeled as `SDSS_QSO`. This subset

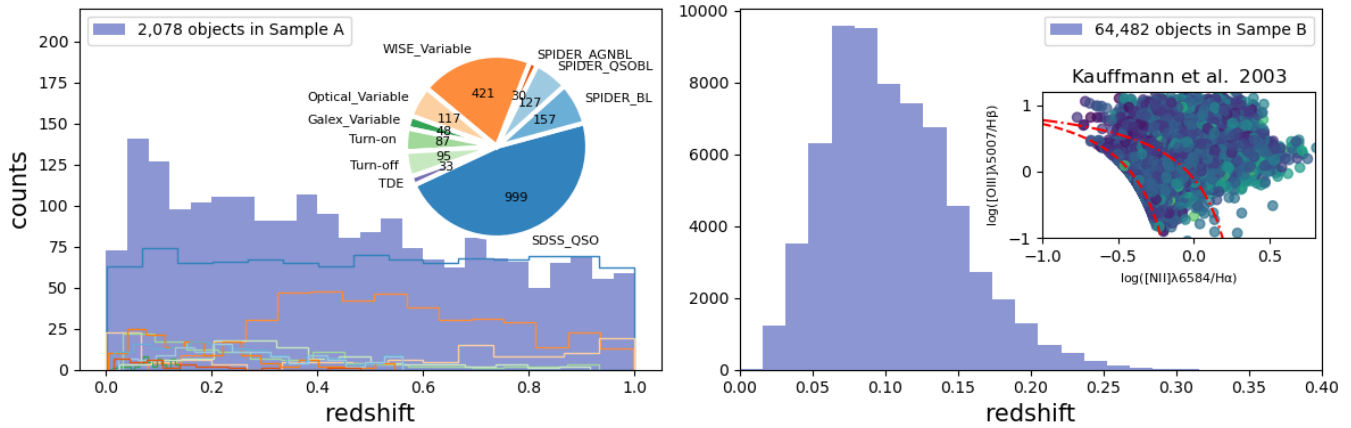


Figure 3. The redshift distribution of the two fully distinct samples is presented here. The left panel corresponds to the first one with approximately 2000 AGNs; the inset pie chart shows the number of objects from each subsample. The right panel comprises approximately 65,000 type-2 AGNs from G. Kauffmann et al. (2003). The inset plot shows the sample on the BPT diagram, color-coded by the mean fractional variation of the light curves. All objects in this subsample sit in the composite and AGN parts of the BPT by construct.

is intended to serve as a broad reference population rather than a physically homogeneous class. The selection is therefore performed without imposing cuts on luminosity, black hole mass, host-galaxy properties, or variability metrics. To prevent this large parent catalog from dominating the learned manifold and to reduce unnecessary computational cost, we construct a stratified random subsample with an approximately flat redshift distribution over $0 < z < 1$. Specifically, the redshift range is divided into equal-width bins, and an equal number of quasars is randomly drawn from each bin, ensuring uniform redshift coverage while preserving diversity in intrinsic properties. Because the SDSS quasar catalog is large, many realizations of this stratified random sampling are possible. We therefore repeated the sampling procedure multiple times using different random draws and verified that the overall structure of the learned manifold and the qualitative relationships discussed below are stable across realizations. This subset intentionally overlaps with other labels in sample A, reflecting the fact that variability-selected AGN populations are not mutually exclusive.

We include AGNs from the SPectroscopic IDentification of ERosita Sources (SPIDERS) survey, which obtained targeted optical spectroscopy for X-ray-selected AGNs (N. Clerc et al. 2016; T. Dwelly et al. 2017; M. Salvato et al. 2018). Of the $\sim 12,600$ successful SPIDERS spectra, roughly half are classified as quasars, ~ 4800 as galaxies, and the remainder as stars. We exclude stellar sources and restrict the sample to $0 < z < 1$. Broad-line AGNs and quasars are labeled as SPIDERS_AGNBL and SPIDERS_QSOBL, respectively, with their union denoted SPIDERS_BL. Narrow-line AGNs are labeled SPIDERS_AGN. This selection retains AGN-like extragalactic sources while excluding stars and purely star-forming systems.

Sample A further includes variability-selected AGNs across multiple wavelengths. In the ultraviolet, we include 48 AGNs identified as variable in the GALEX time-domain survey (S. Gezari et al. 2013) by E. J. Wasleske et al. (2022), labeled GALEX_Variable. Optical variables (Optical_Variable) consist of 117 AGNs identified in the 3 yr COSMOS dataset from the Very Large Telescope Survey Telescope (M. Capaccioli & P. Schipani 2011) by D. De Cicco et al. (2019). In the mid-infrared, we include WISE-variable AGNs

selected from the R90 catalog of R. J. Assef et al. (2018). Of the ~ 4.5 million WISE-detected AGNs in this catalog, A. Prakash et al. (2019) identified 687 highly variable sources, of which 421 fall within our adopted redshift range and are labeled WISE_Variable. No additional cuts are applied to these variability-selected subsamples beyond the original selection criteria and redshift range.

We additionally compile 182 CLAGNs from the literature (S. M. LaMassa et al. 2015; C. L. MacLeod et al. 2016; J. J. Ruan et al. 2016; Q. Yang et al. 2018; M. J. Graham et al. 2020; Z. Sheng et al. 2020; P. J. Green et al. 2022; W. J. Hon et al. 2022; E. López-Navas et al. 2022). These are labeled as 87 turn-on and 95 turn-off CLAGNs, based on the reported direction of their spectral transitions. The selection criteria for CLAGNs vary substantially across studies, ranging from large-amplitude optical variability thresholds (e.g., C. L. MacLeod et al. 2016) to spectroscopically confirmed changes in Balmer line fluxes or widths (e.g., M. J. Graham et al. 2020; P. J. Green et al. 2022). Other works employ mid-infrared variability (Z. Sheng et al. 2020) or ML-assisted photometric selection followed by spectroscopic confirmation (W. J. Hon et al. 2022; E. López-Navas et al. 2022).

Given this heterogeneity, we do not attempt to homogenize CLAGN selection or assess the reliability of individual classifications. Instead, we group all confirmed or high-confidence sources into turn-on and turn-off categories and treat the labels as external information. Our aim is to test whether unsupervised manifold learning applied solely to light-curve morphology can recover or distinguish these labeled transitions in variability space. Objects without clearly defined spectral transitions or ambiguous turn-on/off directionality were excluded.

Finally, we include 33 confirmed TDEs from ZTF (S. van Velzen et al. 2021; E. Hammerstein et al. 2023; J. J. Somalwar et al. 2023). TDEs are included because they exhibit extreme variability and spectroscopic changes that can resemble those seen in CLAGNs, offering a useful comparison class. We note that most TDE searches explicitly exclude known AGNs or sources with AGN-like mid-infrared colors to reduce contamination (e.g., D. Stern et al. 2005; S. van Velzen et al. 2021), implying that TDEs occurring in active galaxies may be underrepresented despite being astrophysically

expected (e.g., C.-H. Chan et al. 2019; J. M. M. Neustadt et al. 2020; C. Ricci et al. 2020; Y. Wang et al. 2024).

4.2. Homogeneous Narrow-line AGNs (Sample B)

Sample B is drawn from the type-2 AGN catalog of G. Kauffmann et al. (2003) and has a narrow redshift distribution centered at $z \sim 0.1$, as shown in the right panel of Figure 3. These AGNs are selected based on their positions above the star-forming sequence in the composite and AGN regions of the Baldwin, Philips, & Terlevich (BPT) diagram (J. A. Baldwin et al. 1981) and have well-characterized host-galaxy and emission-line properties.

This sample provides a complementary counterpart to sample A. Its uniform selection, large size, and consistent spectroscopic measurements enable a direct comparison between the variability manifold and independent physical diagnostics, including [O III] λ 5007 luminosity, stellar mass, $D_n(4000)$, and $H\delta_A$. These quantities are not used in constructing the manifold and therefore offer an external validation of whether photometric variability morphology encodes physically relevant information.

We note that sample B is dominated by relatively low-luminosity AGNs, including Seyferts and LINERs, and does not fully overlap in luminosity with the more powerful quasars in sample A. Extending this analysis to higher-luminosity obscured quasars—for example, using optically selected type-2 quasar samples from SDSS (N. L. Zakamska et al. 2003; R. Reyes et al. 2008)—is a natural avenue for future work and would allow testing whether the variability-manifold structure persists across a broader range of accretion power.

5. Manifold Learning and Dimensionality Reduction

In this work, we use Uniform Manifold Approximation and Projection (UMAP; L. McInnes et al. 2018) as a nonlinear dimensionality reduction technique to construct low-dimensional representations of multiband AGN light curves. UMAP is designed to preserve local neighborhood structure in high-dimensional data while providing a computationally efficient embedding suitable for visualization and exploratory analysis. Throughout this work, UMAP is used to examine similarity structure in multiband time-domain photometric behavior. While the resulting low-dimensional representations are not intended to provide precise metric distances or formal hypothesis tests, they offer an interpretable framework for identifying clustering, gradients, and transitions in variability space.

Briefly, UMAP constructs a weighted k -nearest-neighbor graph from the input data using a chosen distance metric, encoding local similarity relationships between points. This graph is interpreted as a fuzzy topological representation of the data under the assumption that it is sampled from a Riemannian manifold. An optimization procedure then seeks a low-dimensional embedding that preserves these local relationships as faithfully as possible. The resulting projection emphasizes relative similarity structure rather than absolute metric distances, and is therefore well suited for comparing multiband time-domain variability patterns across heterogeneous datasets.

When configuring UMAP, three parameters play a primary role: the number of neighbors, the minimum distance, and the distance metric. The number of neighbors controls the scale

over which local structure is defined. Larger values incorporate information from broader neighborhoods and tend to emphasize more global trends in the data, while smaller values prioritize fine-scale structure. In this work, we adopt relatively large neighborhood sizes to suppress stochastic noise and focus on broad variability similarities rather than object-level idiosyncrasies. The minimum distance parameter governs how tightly points are allowed to cluster in the low-dimensional space; we use a comparatively large value to discourage overly compact clusters and instead highlight continuous distributions and separations between populations.

The choice of distance metric is particularly important because it determines how similarity between light curves is quantified. While Euclidean distance is commonly used and performs well for uniformly sampled, low-dimensional data, it is less well-suited for irregular, asynchronous time series. Manhattan distance can mitigate some high-dimensional effects, but still treats time series as static vectors. For this reason, we also explore dynamic time warping (DTW), a distance metric specifically designed for time-series comparison.

DTW computes the similarity between two sequences by identifying an optimal alignment that allows local stretching or compression in time, thereby accommodating phase offsets and irregular sampling. This flexibility makes DTW well-suited for AGN light curves, where variability timescales, cadences, and phase relationships can differ significantly across bands and objects. DTW has been widely used in temporal pattern recognition across many domains, including speech processing, gesture recognition, and biological signal analysis (e.g., E. Keogh & C. A. Ratanamahatana 2005). Its main limitation is computational cost, particularly for multivariate time series, since it requires evaluation of a full alignment cost matrix.

To assess the impact of preprocessing and distance choices, Figure 4 presents UMAP projections for sample A using three distance metrics (Euclidean, Manhattan, and DTW) applied to light curves interpolated via both nearest-neighbor linear interpolation and GP regression. These projections are colored by the arctangent of the excess variance, a scalar summary of mean fractional variability amplitude. The arctangent transformation maps the nonnegative excess variance values onto a bounded range, reducing the influence of extreme outliers while preserving relative ordering. Dark blue corresponds to lower variability amplitude and yellow to higher variability amplitude.

Among the tested configurations, DTW applied to GP-regressed light curves yields one of the clearest embeddings in the sense that sources with similar variability amplitudes populate contiguous regions of the projection and form smooth gradients. Other combinations—such as GP regression with Euclidean distance—also recover interpretable structure and are not excluded from consideration. We therefore emphasize that DTW+GP is not uniquely optimal, but represents one effective choice for the subsequent analysis. Figure 4 is included to demonstrate the sensitivity of the embedding to preprocessing and distance choices, rather than to define a single preferred configuration.

To aid interpretation of the embeddings, we color-code UMAP projections using three scalar quantities: redshift, mean brightness, and mean fractional variability. This allows us to assess whether the learned manifold organizes sources in a

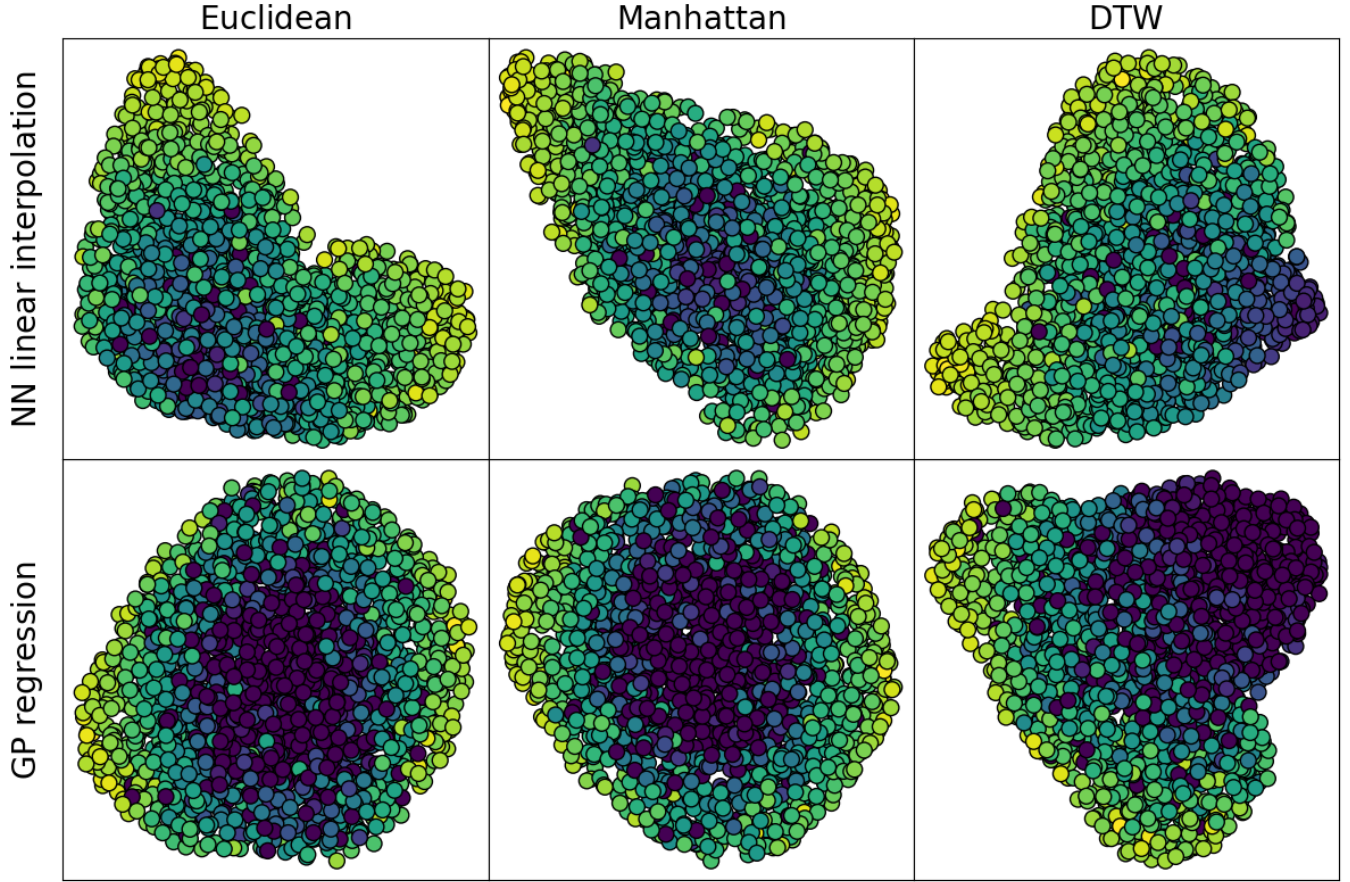


Figure 4. UMAP projections of the WISE W1 light curves for sample A, generated using different combinations of distance metrics (columns) and interpolation methods (rows). This sample includes AGNs compiled from archival variability catalogs, including changing-look AGNs (CLAGNs) and tidal disruption events (TDEs). Each point represents a source, color-coded by the mean fractional variation of its light curve. The color scale ranges from dark blue (low variability) to yellow (high variability). These projections allow us to assess how different preprocessing choices affect the learned manifold structure. Among the configurations, Gaussian process (GP) interpolation combined with the dynamic time warping (DTW) distance metric produces the most coherent structure, with a clear gradient in fractional variation. The presence of this gradient indicates that variability amplitude is a key factor shaping the manifold, suggesting that this configuration best preserves physically meaningful variability information.

manner consistent with known observational properties, without using these quantities as inputs to the embedding.

The brightness parameter shown in the UMAP plots is defined as

$$F_{\text{mean}} = \log \left(\sum_{b=1}^M \langle F_b \rangle_{\sigma} \right), \quad (1)$$

where the sum runs over the M photometric bands, F_b denotes the flux values in the b th band, and $\langle F_b \rangle_{\sigma}$ is the time-averaged flux after applying a 5σ clipping to suppress outliers.

The mean fractional variability is defined as

$$\text{Fvar}_{\text{mean}} = \arctan \left(\sum_{b=1}^M \frac{\sqrt{\sigma_b^2 - \delta_b^2}}{\langle F_b \rangle} \right), \quad (2)$$

where

$$\sigma_b^2 = \frac{1}{N} \sum_{i=1}^N (F_i - \langle F \rangle)^2, \quad (3)$$

$$\delta_b^2 = \frac{1}{N} \sum_{i=1}^N (\Delta F_i)^2, \quad (4)$$

and ΔF_i denotes the measurement uncertainty. This formulation accounts for observational noise following the standard excess-variance definition (B. M. Peterson 2001). When the estimated intrinsic variance is smaller than the mean squared uncertainty, the contribution is set to zero, ensuring that $\text{Fvar}_{\text{mean}}$ remains nonnegative. The logarithmic and arctangent transformations are applied solely for visualization purposes, to compress dynamic range and enhance interpretability.

6. Results

Figure 5 shows the UMAP projections learned using the WISE W1 band alone. As expected, given the flux normalization applied during preprocessing, no systematic trend with mean brightness is visible in the left panel. Likewise, the absence of a redshift gradient in the middle panel indicates that the manifold is primarily structured by patterns in the time-domain light curves rather than by absolute flux scale or rest-frame wavelength. Although the WISE W1 band probes rest-frame wavelengths spanning approximately $2.8\text{--}1.7 \mu\text{m}$ across sample A, this variation does not appear to dominate the learned variability structure. In contrast, a clear gradient in mean fractional variability is visible in the right panel,

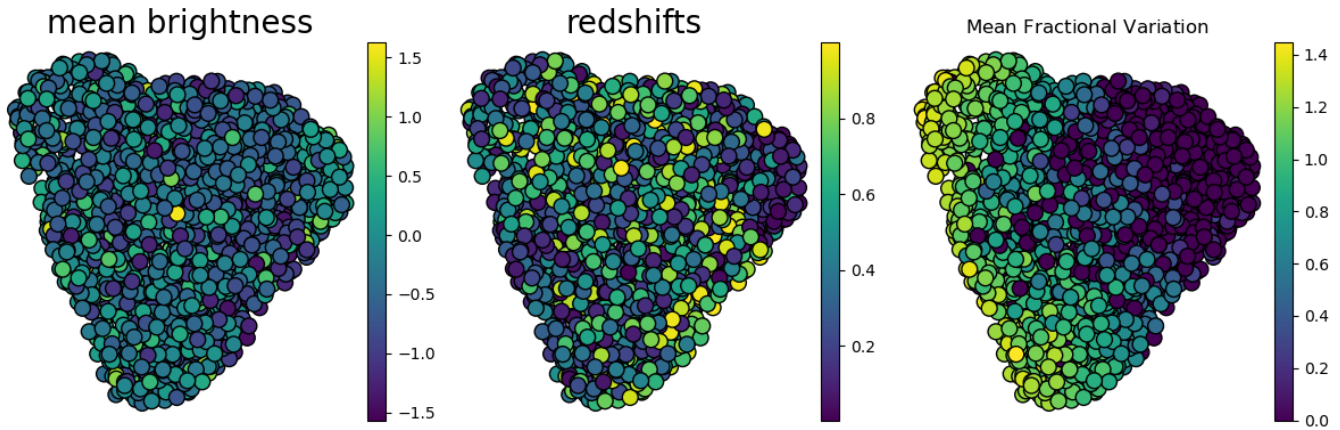


Figure 5. This figure presents UMAP projections of the WISE W1 light curves for the first AGN sample, using GP regression and the DTW distance metric (as in the bottom-right panel of Figure 4). Points are color-coded by mean brightness, redshift, and mean fractional variation (left to right). As expected, given that the manifold was constructed from normalized light curves, there is no clear structure associated with redshift or brightness. In contrast, a pronounced gradient is observed with mean fractional variation, indicating that variability amplitude plays a dominant role in shaping the learned manifold. This confirms that the embedding captures physically meaningful differences in AGN variability behavior.

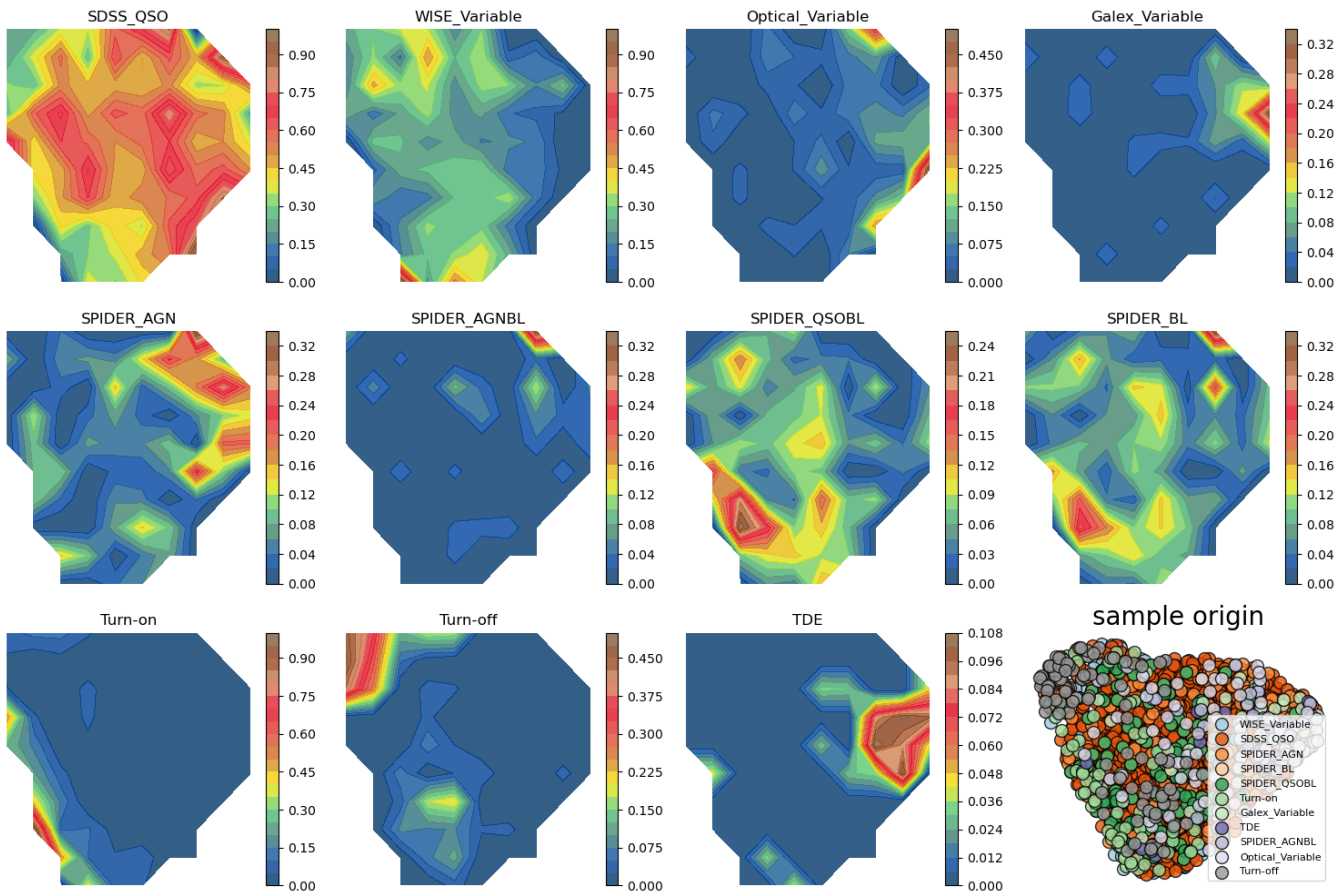


Figure 6. Separation of different labels in the first sample is illustrated using UMAPs, with parameters similar to those in Figure 5. The bottom right panel displays the origin labels for all objects in this sample. Each panel shows a 2D histogram over the UMAP manifold for a particular AGN class label. The color bar indicates the local point density for that class in the UMAP embedding space, with lighter shades representing regions of higher density. This allows for a visual comparison of where each labeled population is concentrated or dispersed within the learned manifold. The degree of separation between these distributions indicates how well the manifold distinguishes between different AGN classes based on variability. These projections use the same preprocessing and UMAP parameters as in Figure 5, namely GP-regressed WISE W1 light curves and the DTW distance metric.

indicating that variability amplitude is one of the dominant organizing axes in this single-band embedding.

To examine how known AGN populations project onto this learned time-domain variability space, Figure 6 shows the

spatial distributions of sources with different external labels overlaid on the same UMAP embedding. The SDSS quasar subset spans much of the manifold, consistent with its heterogeneous variability behavior. In contrast, the WISE-

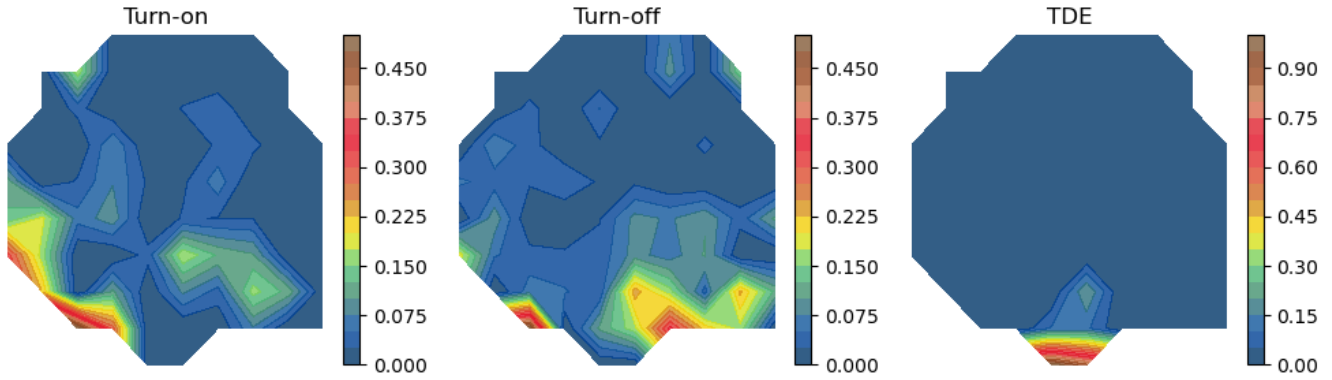


Figure 7. Similar to Figure 6, but trained on the ZTF g band alone to show the similarity of TDEs (right panel) to CLAGNs (two left panels) in optical.

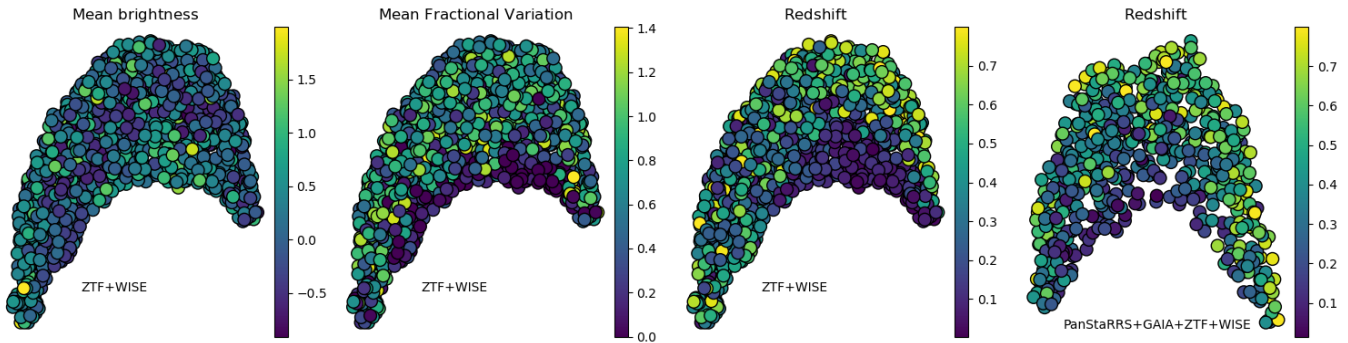


Figure 8. Similar to Figure 5, with the three left panels generated using the combined ZTF+WISE bands and the right panel from Pan-STARRS+Gaia+ZTF+WISE. With the increased number of wave bands covering a broader range of the spectral energy distribution, a redshift trend becomes apparent, and the trend with mean fractional variation is less prominent.

variable AGNs preferentially populate the left-hand side of the embedding, while UV- and optically selected variables occupy overlapping but distinct regions. This pattern suggests that variability selected at different wavelengths emphasizes different aspects of the underlying time-domain behavior, although we do not assign statistical significance to these visual trends.

The middle row of Figure 6 shows the SPIDERS subsamples. Differences between broad-line and narrow-line objects are subtle in this representation, but a qualitative tendency is visible: sources without broad emission lines preferentially occupy regions associated with lower mean fractional variability, while broad-line quasars extend into higher-variability regions. These trends are descriptive and should be interpreted cautiously, particularly given the heterogeneous selection of the SPIDERS sample.

The bottom row highlights CLAGNs and TDEs. The turn-on and turn-off CLAGNs occupy overlapping but distinct regions, particularly at higher variability amplitudes. TDEs, by contrast, cluster toward a separate edge of the manifold and do not strongly overlap with the CLAGN populations in the WISE W1-only embedding. This behavior is consistent with the extreme, transient nature of TDE variability, but the small sample size precludes firm conclusions.

The partial overlap between turn-on CLAGNs and broad-line SPIDERS quasars is consistent with previous estimates that a nonnegligible fraction of highly variable quasars may undergo changing-look transitions, depending on monitoring

cadence and timescale (C. L. MacLeod et al. 2016, 2019). This overlap highlights the continuum between strong quasar variability and spectroscopically confirmed changing-look behavior.

When repeating the analysis using only a single ZTF band (the g band; Figure 7), the distinction between turn-on and turn-off CLAGNs becomes less apparent, and TDEs display more similar time-domain behavior. Given the substantially shorter observational baseline of ZTF compared to WISE, this result is consistent with the interpretation that variability timescale is an important discriminator between these populations (W. J. Hon et al. 2022). However, because most TDE searches explicitly exclude AGNs to minimize contamination, the true degree of overlap between TDEs and extreme AGN variability remains uncertain. Continued long-term monitoring of turn-on CLAGNs may help clarify this relationship.

Rather than learning the manifold from a single band, we can combine observations across facilities. Figures 8 and 9 show UMAP embeddings constructed from three ZTF bands together with the two WISE bands. This combination leverages the complementary strengths of the surveys: ZTF provides high-cadence sampling sensitive to short-timescale variability, while WISE provides a long temporal baseline that captures slower variability modes.

Including Pan-STARRS and Gaia data (right panel of Figure 8) primarily reduces the sample size, as not all objects are detected in these surveys. Because the optical wavelength

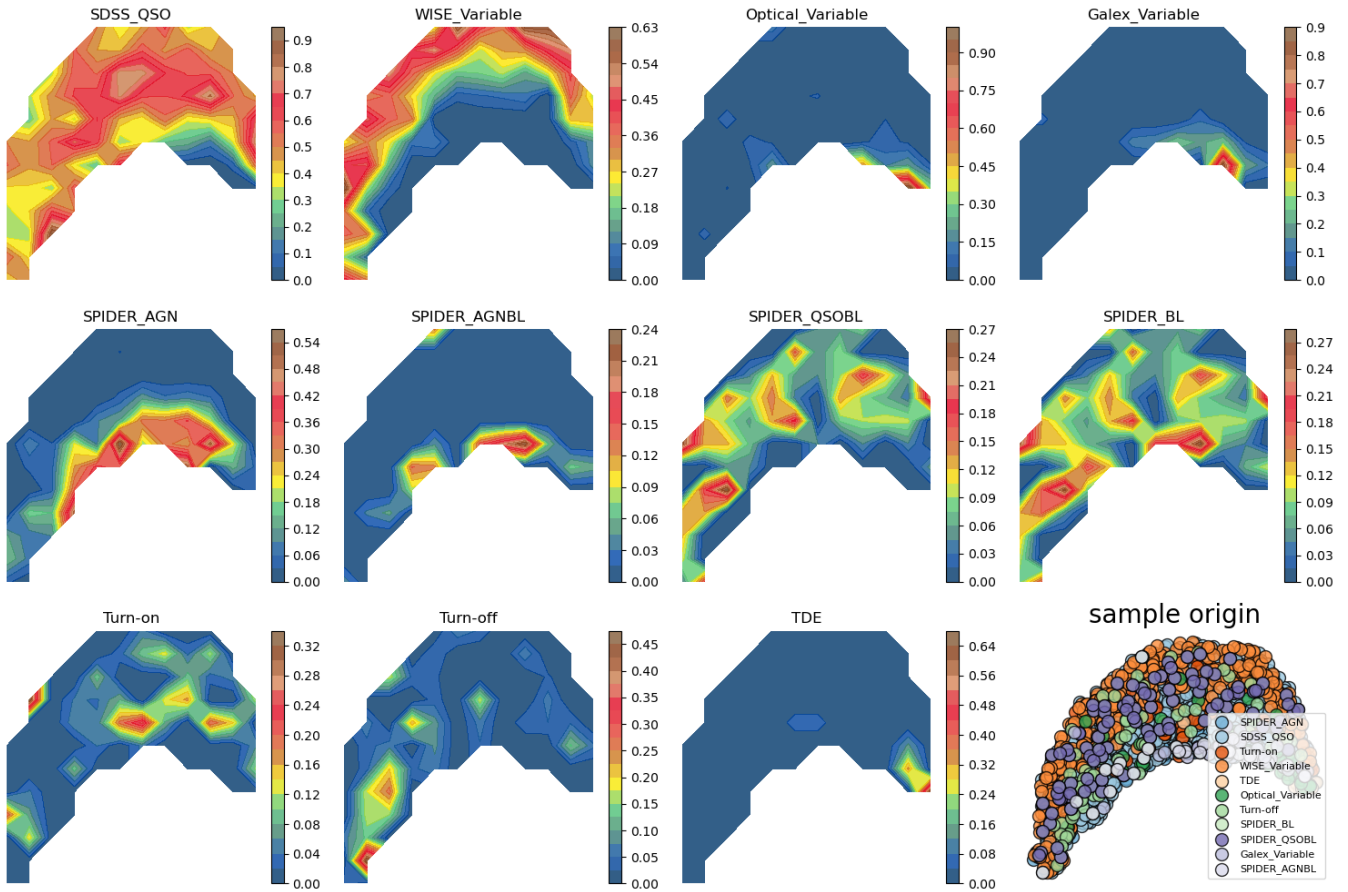


Figure 9. Separation of different labels in sample A is illustrated using UMAPs, with parameters similar to those in Figure 8. The bottom right panel displays the origin labels for all objects in this sample. The remaining panels feature 2D histograms, each representing the distribution of individual labels.

coverage overlaps substantially with ZTF, this inclusion does not add significant new information to the manifold in practice, although the framework places no formal limitation on incorporating additional bands.

The combined ZTF+WISE manifold spans a broader portion of the spectral energy distribution and now exhibits a mild redshift gradient, reflecting the inclusion of multiple wavelength regimes. Variability amplitude is less dominant than in the single-band case, but still contributes to the overall structure. As shown in Figure 9, UV-, optical-, and WISE-variable AGNs continue to populate partially distinct regions, though these patterns are influenced by redshift and sample composition. The SPIDERS sources again span much of the embedding, with narrow-line objects preferentially occupying lower-variability regions. The CLAGN populations are less distinctly separated in this multiband space, likely reflecting both redshift effects and the increased dimensional complexity of the combined data. TDEs, however, remain confined to a relatively compact region at the edge of the manifold.

Figure 10 presents the ZTF+WISE manifold for sample B, the homogeneous type-2 AGN sample drawn from G. Kauffmann et al. (2003). Unlike sample A, this dataset provides uniform spectroscopic measurements, enabling a direct comparison between patterns in the learned time-domain space and independent physical diagnostics. Despite the narrow redshift range of this sample, the learned manifold

exhibits clear gradients when colored by [O III] luminosity, $H\delta_A$, $D_n(4000)$, stellar mass, and mean fractional variability—none of which were used in constructing the embedding.

The projection reveals that [O III] luminosity and $H\delta_A$ absorption increase toward the central regions of the manifold, while the 4000 Å break weakens, indicating younger stellar populations. The mean fractional variability measured from the light curves also becomes more pronounced toward these regions. Stellar mass shows a distinct trend approximately perpendicular to these gradients, possibly reflecting differences in host-galaxy properties or dust attenuation. These orthogonal trends suggest that the manifold captures multiple, partially independent axes of structure present in the high-dimensional time-domain data.

Importantly, we do not find a strong correspondence between position on the BPT diagram and location in the variability manifold, indicating that the embedding is not simply reproducing traditional spectroscopic classifications. Instead, the manifold organizes sources according to similarities in their multiband time-domain photometric behavior, with spectroscopic properties varying smoothly across this space. This behavior mirrors known correlations between stellar mass, stellar age, and [O III] luminosity in type-2 AGNs (G. Kauffmann et al. 2003), and demonstrates that photometric time-domain data alone encode information linked to these independently measured properties.

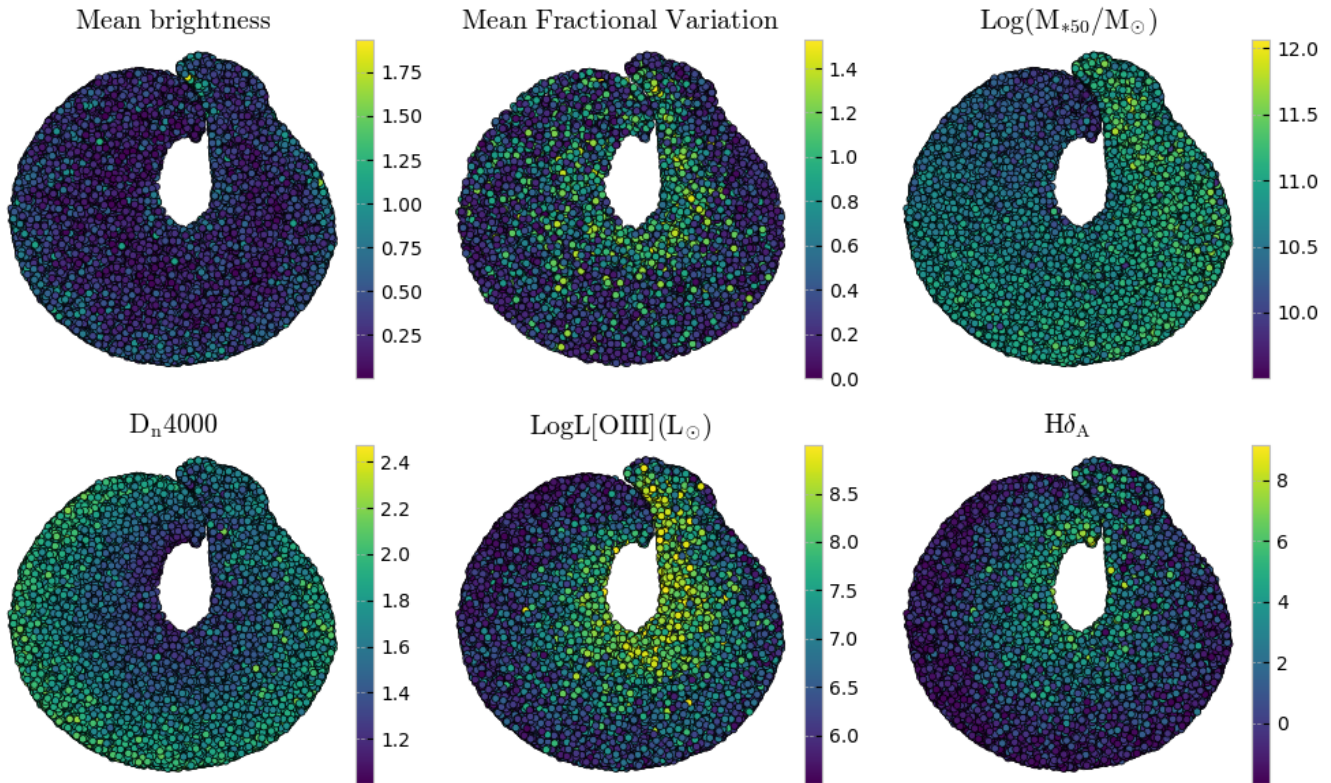


Figure 10. This figure presents the UMAPs trained on the GP-regressed ZTF+WISE band light curves of type-2 AGNs in sample B. The maps from top-left to bottom-right are color-coded by the mean brightness and mean fractional variation (measured from the light curves) and the central stellar mass ($\text{Log}(M_{*50}/M_{\odot})$), D_n4000 , [O III] corrected luminosity, and $H\delta_A$ (directly from G. Kauffmann et al. 2003). Sample B was used here due to its uniform spectroscopic coverage, allowing us to compare the manifold structure learned from light curves with independent physical tracers of AGN activity.

7. Summary and Discussion

In this study, we explored the use of archival multi-wavelength time-series data to construct a unified representation of AGN variability using manifold learning. By integrating heterogeneous light curves from multiple surveys and applying nonlinear dimensionality reduction, we show that observed time-domain variability can be organized into a low-dimensional space that captures coherent structure associated with external AGN labels and with independently measured spectroscopic properties.

The purpose of this work is to characterize similarities and differences in observed multiband variability behavior, not to model or identify the physical mechanisms that drive AGN variability. Instead, we present a data-driven, unsupervised framework for comparing time-domain behavior across diverse samples and surveys. Using photometric light curves alone, without labels or spectroscopic inputs, we construct high-dimensional representations of AGN variability and examine their projections for two fully distinct samples: a heterogeneous compilation of variability-selected AGNs (sample A) and a large, homogeneous sample of type-2 AGNs with uniform spectroscopic measurements (sample B).

The main results can be summarized as follows:

1. Without using class labels during training, the learned manifolds organize variability-selected AGN populations into overlapping but nonidentical regions. In sample A, WISE-variable AGNs cluster more tightly than UV- or

optically selected variables, consistent with the longer variability timescales probed in the mid-infrared.

2. Turn-on and turn-off CLAGNs occupy partially distinct regions of the manifold, indicating systematic differences in their multiband time-domain behavior that emerge without supervision.
3. TDEs populate restricted regions of the variability space, particularly distinct in mid-infrared bands, while showing increased overlap with CLAGNs in optical-only representations.
4. In the homogeneous type-2 AGN sample (sample B), manifold coordinates vary smoothly with independently measured spectroscopic and host-galaxy properties, including [O III] luminosity, $D_n(4000)$, $H\delta_A$, stellar mass, and mean fractional variability, despite none of these quantities being used in constructing the manifold.

Taken together, the results from sample A demonstrate that AGN populations selected using different variability criteria occupy overlapping but structured regions of variability space. Differences between mid-infrared, optical, and UV variability-selected AGNs reflect wavelength-dependent sensitivity to variability amplitude and timescale, with WISE-selected sources exhibiting more compact distributions likely driven by long-term variability. Turn-on and turn-off CLAGNs likewise show systematic offsets, while TDEs cluster in restricted regions despite not being explicitly modeled. These patterns indicate that time-domain variability alone is sufficient to recover characteristic multiband variability signatures at a descriptive level, without requiring labels or

physical inputs, though we do not assess statistical separability or define sharp population boundaries.

The clearest indication that the learned manifold captures information beyond sample labeling comes from sample B, which serves as an internal validation set. In this homogeneous type-2 AGN sample, manifold coordinates correlate smoothly with spectroscopic diagnostics that trace nuclear activity and host-galaxy properties. We use the term “physically meaningful” in a restricted sense: the variability manifold learned from photometric time series alone captures structure that is correlated with independently measured physical quantities, without implying a causal relationship or identifying the underlying drivers of variability. This demonstrates that similarities in observed variability behavior encode information related to physical properties measured through spectroscopy.

Within the learned manifold, two broad types of structure are evident. First, sources sharing external labels often occupy similar regions, reflecting common variability behavior. Second, continuous gradients appear that correlate with spectroscopic observables, particularly in sample B. These structures arise solely from the time-domain data and were not imposed through feature engineering or supervised learning. At the same time, spatial proximity in the manifold reflects similarity in observed variability behavior and should not be interpreted as evidence for shared physical mechanisms or evolutionary pathways. Apparent overlap between populations often reflects similarities in variability amplitude or timescale that may arise from different underlying processes. The manifold is therefore best viewed as an exploratory diagnostic that complements, rather than replaces, traditional physical measurements.

Several limitations and directions for future work follow naturally from this analysis. First, the external labels used here are heterogeneous and not standardized across studies, particularly for CLAGNs, which are identified using diverse selection criteria and confidence thresholds. While our goal is not to validate individual classifications, this heterogeneity highlights the need for more uniform definitions in future time-domain studies.

Second, although we allow for non-Euclidean distance metrics in constructing the variability space, the use of UMAP assumes that the data lie on a Riemannian manifold. This assumption may not strictly hold for light curves, which can exhibit nonstationary behavior, multiple characteristic timescales, and more complex geometric or topological structure. Future work will explore alternative representations and quantitative diagnostics, such as neighborhood purity or completeness, to characterize structure in variability space more rigorously.

Third, the preprocessing steps required to unify asynchronous, multiband light curves involve trade-offs. Temporal alignment simplifies analysis but may obscure physically meaningful time delays between bands. Differences between optical and infrared variability, for example, can carry important physical information that is not explicitly modeled here. Incorporating relative timing or lag-aware representations is a natural extension of this framework.

The quality and uniformity of archival photometry also remain important considerations. While manifold learning can mitigate some nonsystematic uncertainties, improvements in light-curve extraction, such as consistent forced photometry, point-spread-function-aware measurements, or denoising and

resolution enhancement, would directly benefit analyses of this kind. Community-wide efforts toward standardized photometric pipelines would significantly enhance the interpretability of large time-domain datasets.

More broadly, the learned variability space provides a compact, survey-agnostic representation of time-domain behavior that can be used to compare heterogeneous samples directly. In settings where spectroscopy is unavailable or incomplete, proximity in variability space may serve as a prior for identifying sources likely to share similar physical properties. Rare populations such as CLAGNs and TDEs occupy restricted regions of the manifold, enabling the identification and prioritization of new candidates for spectroscopic follow-up based solely on their multiband variability behavior. A quantitative assessment of targeting efficiency represents a natural next step beyond the proof-of-concept demonstration presented here.

While our analysis focuses on AGNs, the methodology is broadly applicable to other classes of variable and transient sources. Upcoming surveys such as LSST and the Roman Space Telescope will produce vast, heterogeneous time-domain datasets for which unsupervised, time-domain representations will be increasingly valuable.

Finally, we emphasize the exploratory nature of this work. We do not assess the statistical significance of the clustering patterns observed in the UMAP projections, which are used primarily for visualization and hypothesis generation. Moreover, while UMAP preserves local structure effectively, it may distort global relationships between clusters (D. Kobak & P. Berens 2019; D. Kobak & G. C. Linderman 2021; Y. Wang et al. 2021; V. Marx 2024). Distances between widely separated regions in the embedding should therefore be interpreted with caution. Future work will focus on developing quantitative diagnostics for structure in variability space and applying this framework to larger and more diverse samples.










Acknowledgments

This work made extensive use of the Fornax Science Console, part of the NASA Astrophysics cloud-based Fornax Initiative jointly developed by Goddard Space Flight Center’s Astrophysics Projects Division and the HEASARC, IRSA, and MAST archives. S.H. thanks R. Chary for thoughtful discussions and feedback. We thank the referee for constructive comments that improved the clarity and scope of this work.

Software: Astropy (Astropy Collaboration et al. 2018), ChatGPT (OpenAI 2024—for language editing and readability improvements), HEALPix (K. M. Górski et al. 2005), scikit-learn (F. Pedregosa et al. 2011), UMAP (L. McInnes et al. 2018).

Funding Acknowledgment: The research was carried out in part at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (80NM0018D0004).

ORCID iDs

Shoubaneh Hemmati  <https://orcid.org/0000-0003-2226-5395>
 Jessica Krick  <https://orcid.org/0000-0002-2413-5976>
 Daniel Stern  <https://orcid.org/0000-0003-2686-9241>
 Vandana Desai  <https://orcid.org/0000-0002-1340-0543>
 Andreas Faisst  <https://orcid.org/0000-0002-9382-9832>
 Lucas Martín-García  <https://orcid.org/0009-0009-7840-931X>
 Varoujan Gorjian  <https://orcid.org/0000-0002-8990-2101>
 Aryana Haghjoo  <https://orcid.org/0009-0006-3071-7143>
 Farnik Nikakhtar  <https://orcid.org/0000-0002-3641-4366>

Troy Raen  <https://orcid.org/0000-0002-3031-5279>
 Sogol Sanjaripour  <https://orcid.org/0009-0009-3048-9090>
 Brigitta M. Sipőcz  <https://orcid.org/0000-0002-3713-6337>
 David Shupe  <https://orcid.org/0000-0003-4401-0430>

References

- Antonucci, R. 1993, *ARA&A*, 31, 473
 Antonucci, R. R. J. 1984, *ApJ*, 278, 499
 Assef, R. J., Stern, D., Noirot, G., et al. 2018, *ApJS*, 234, 23
 Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, *AJ*, 156, 123
 Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, *PASP*, 93, 5
 Beck, W., Bokde, N., Asencio-Cortés, G., & Kulat, K. 2018, *The R Journal*, 10, 218
 Bellm, E. C., Kulkarni, S. R., Graham, M. J., et al. 2019, *PASP*, 131, 018002
 Bentz, M. C., Denney, K. D., Grier, C. J., et al. 2013, *ApJ*, 767, 149
 Burke, C. J., Shen, Y., Blaes, O., et al. 2021, *Sci*, 373, 789
 Cackett, E. M., Bentz, M. C., & Kara, E. 2021, *iSci*, 24, 102557
 Capaccioli, M., & Schipani, P. 2011, *Msngr*, 146, 2
 Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, arXiv:1612.05560
 Chan, C.-H., Piran, T., Krolik, J. H., & Saban, D. 2019, *ApJ*, 881, 113
 Clerc, N., Merloni, A., Zhang, Y. Y., et al. 2016, *MNRAS*, 463, 4490
 De Cicco, D., Paolillo, M., Falocco, S., et al. 2019, *A&A*, 627, A33
 Dwelly, T., Salvato, M., Merloni, A., et al. 2017, *MNRAS*, 469, 1065
 Elorrieta, F., Eyheramendy, S., Palma, W., & Ojeda, C. 2021, *MNRAS*, 505, 1105
 Faisst, A. L., Prakash, A., Capak, P. L., & Lee, B. 2019, *ApJL*, 881, L9
 Flewelling, H. A., Magnier, E. A., Chambers, K. C., et al. 2020, *ApJS*, 251, 7
 Folgado, D., Barandas, M., Matias, R., et al. 2018, *PatRe*, 81, 268
 Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, *A&A*, 595, A1
 Gaia Collaboration, Vallenari, A., Brown, A. G. A., et al. 2023, *A&A*, 674, A1
 Gezari, S., Martin, D. C., Forster, K., et al. 2013, *ApJ*, 766, 60
 Gilbertson, C., Ford, E. B., Jones, D. E., & Stenning, D. C. 2020, *ApJ*, 905, 155
 Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, *ApJ*, 622, 759
 Graham, M. J., Ross, N. P., Stern, D., et al. 2020, *MNRAS*, 491, 4925
 Green, P. J., Pulgarin-Duque, L., Anderson, S. F., et al. 2022, *ApJ*, 933, 180
 Hammerstein, E., van Velzen, S., Gezari, S., et al. 2023, *ApJ*, 942, 9
 Hemmati, S., Capak, P., Masters, D., et al. 2019a, *ApJ*, 877, 117
 Hemmati, S., Capak, P., Pourrahmani, M., et al. 2019b, *ApJL*, 881, L14
 Hon, W. J., Wolf, C., Onken, C. A., Webster, R., & Auchettl, K. 2022, *MNRAS*, 511, 54
 Hu, Z., & Tak, H. 2020, *AJ*, 160, 265
 Kauffmann, G., Heckman, T. M., Tremonti, C., et al. 2003, *MNRAS*, 346, 1055
 Kelly, B. C., Bechtold, J., & Siemiginowska, A. 2009, *ApJ*, 698, 895
 Keogh, E., & Ratanamahatana, C. A. 2005, *KIS*, 7, 358
 Kobak, D., & Berens, P. 2019, *NatCo*, 10, 5416
 Kobak, D., & Linderman, G. C. 2021, *NatBi*, 39, 156
 Koshida, S., Minezaki, T., Yoshii, Y., et al. 2014, *ApJ*, 788, 159
 LaMassa, S. M., Cales, S., Moran, E. C., et al. 2015, *ApJ*, 800, 144
 López-Navas, E., Martínez-Aldama, M. L., Bernal, S., et al. 2022, *MNRAS*, 513, L57
 Lyke, B. W., Higley, A. N., McLane, J. N., et al. 2020, *ApJS*, 250, 8
 MacLeod, C. L., Green, P. J., Anderson, S. F., et al. 2019, *ApJ*, 874, 8
 MacLeod, C. L., Ivezić, Z., Kochanek, C. S., et al. 2010, *ApJ*, 721, 1014
 MacLeod, C. L., Ross, N. P., Lawrence, A., et al. 2016, *MNRAS*, 457, 389
 Mainzer, A., Bauer, J., Grav, T., et al. 2011, *ApJ*, 731, 53
 Marx, V. 2024, *NatMe*, 21, 930
 Masci, F. J., Laher, R. R., Rusholme, B., et al. 2019, *PASP*, 131, 018003
 Masters, D., Capak, P., Stern, D., et al. 2015, *ApJ*, 813, 53
 McHardy, I. M., Koerding, E., Knigge, C., Uttley, P., & Fender, R. P. 2006, *Natur*, 444, 730
 McInnes, L., Healy, J., & Melville, J. 2018, arXiv:1802.03426
 Meisner, A. M., Caselden, D., Schlafly, E. F., & Kiwy, F. 2023, *AJ*, 165, 36
 Moritz, S., Sardá, A., Bartz-Beielstein, T., Zaefferer, M., & Stork, J. 2015, arXiv:1510.03924
 Netzer, H. 2015, *The Physics and Evolution of Active Galactic Nuclei* (Cambridge Univ. Press)
 Neustadt, J. M. M., Holoien, T. W. S., Kochanek, C. S., et al. 2020, *MNRAS*, 494, 2538
 OpenAI 2024, ChatGPT, Large language model, <https://chat.openai.com/>
 Pantoja, R., Catelan, M., Pichara, K., & Protopapas, P. 2022, *MNRAS*, 517, 3660
 Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *JMLR*, 12, 2825
 Peterson, B. M. 2001, in *Advanced Lectures on the Starburst-AGN Connection*, ed. I. Aretxaga, D. Kunth, & R. Mújica (World Scientific), 3
 Prakash, A., Chary, R. R., Helou, G., et al. 2019, *ApJ*, 883, 154
 Reyes, R., Zakamska, N. L., Strauss, M. A., et al. 2008, *AJ*, 136, 2373
 Ricci, C., Kara, E., Loewenstein, M., et al. 2020, *ApJL*, 898, L1
 Ricci, C., & Trakhtenbrot, B. 2023, *NatAs*, 7, 1282
 Ross, N. P., Graham, M. J., Calderone, G., et al. 2020, *MNRAS*, 498, 2339
 Ruan, J. J., Anderson, S. F., Cales, S. L., et al. 2016, *ApJ*, 826, 188
 Salvato, M., Buchner, J., Budavári, T., et al. 2018, *MNRAS*, 473, 4937
 Sánchez-Sáez, P., Lira, H., Martí, L., et al. 2021, *AJ*, 162, 206
 Sanjaripour, S., Aravindan, A., Canalizo, G., et al. 2025, *ApJ*, 992, 138
 Sanjaripour, S., Hemmati, S., Mobasher, B., et al. 2024, *ApJ*, 977, 202
 Shappee, B. J., Prieto, J. L., Grupe, D., et al. 2014, *ApJ*, 788, 48
 Sheng, Z., Wang, T., Jiang, N., et al. 2020, *ApJ*, 889, 46
 Sokolovsky, K. V., Gavras, P., Karamelas, A., et al. 2017, *MNRAS*, 464, 274
 Somalwar, J. J., Ravi, V., Dong, D. Z., et al. 2023, *ApJ*, 945, 142
 Stern, D., Eisenhardt, P., Gorjian, V., et al. 2005, *ApJ*, 631, 163
 Stern, D., McKernan, B., Graham, M. J., et al. 2018, *ApJ*, 864, 27
 Ulrich, M.-H., Maraschi, L., & Urry, C. M. 1997, *ARA&A*, 35, 445
 Urry, C. M., & Padovani, P. 1995, *PASP*, 107, 803
 van Velzen, S., Gezari, S., Hammerstein, E., et al. 2021, *ApJ*, 908, 4
 Wang, S., Woo, J.-H., Gallo, E., et al. 2025, *ApJ*, 981, 129
 Wang, Y., Graham, M. J., Ford, K. E. S., et al. 2024, arXiv:2406.12096
 Wang, Y., Huang, H., Rudin, C., & Shaposhnik, Y. 2021, arXiv:2012.04456
 Wasleske, E. J., Baldassare, V. F., & Carroll, C. M. 2022, *ApJ*, 933, 37
 Williams, C., & Rasmussen, C. 1995, in *Advances in Neural Information Processing Systems*, ed. D. Touretzky, M. Mozer, & M. Hasselmo, Vol. 8 (MIT Press) https://proceedings.neurips.cc/paper_files/paper/1995/file/7cce53cf9057744271720a370c3c723-Paper.pdf
 Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, *AJ*, 140, 1868
 Yang, Q., Green, P. J., Wu, X.-B., et al. 2025, *ApJ*, 980, 91
 Yang, Q., Wu, X.-B., Fan, X., et al. 2018, *ApJ*, 862, 109
 Yang, Q., Green, P. J., MacLeod, C. L., et al. 2023, *ApJ*, 953, 61
 Zakamska, N. L., Strauss, M. A., Krolik, J. H., et al. 2003, *AJ*, 126, 2125