

# CAZ catalog and optical light curves of 7918 blazar-selected active galactic nuclei

Pouya M. Kouch<sup>1,2,3,\*</sup>, Elina Lindfors<sup>1,2</sup>, Talvikki Hovatta<sup>2,3,4</sup>, Ioannis Liodakis<sup>5</sup>, Karri I. I. Koljonen<sup>6</sup>, Alessandro Paggi<sup>5</sup>, Kari Nilsson<sup>2</sup>, Jenni Jormanainen<sup>1,2</sup>, Vandad Fallah Ramazani<sup>2</sup>, Sofia Kankkunen<sup>3,4</sup>, Folkert Wierda<sup>1,2</sup>, Sarah M. Wagner<sup>7</sup>, and Matthew J. Graham<sup>8</sup>

<sup>1</sup> Department of Physics and Astronomy, University of Turku, 20014 Turku, Finland

<sup>2</sup> Finnish Centre for Astronomy with ESO (FINCA), Quantum, Vesilinnantie 5, University of Turku, 20014 Turku, Finland

<sup>3</sup> Aalto University Metsähovi Radio Observatory, Metsähovintie 114, 02540 Kylmälä, Finland

<sup>4</sup> Aalto University Department of Electronics and Nanoengineering, PL 15500, 00076 Espoo, Finland

<sup>5</sup> Institute of Astrophysics, Foundation for Research and Technology-Hellas, 71110 Heraklion, Greece

<sup>6</sup> Institutt for Fysikk, Norwegian University of Science and Technology, Høgskloreringen 5, Trondheim 7491, Norway

<sup>7</sup> Institut für Theoretische Physik and Astrophysik, Universität Würzburg, 97074 Würzburg, Germany

<sup>8</sup> Division of Physics, Mathematics and Astronomy, California Institute of Technology, Pasadena, CA91125, USA

Received 7 October 2025 / Accepted 6 March 2026

## ABSTRACT

Active galactic nuclei (AGN) are some of the brightest and most variable objects in the Universe. Those with relativistic jets observed at small viewing angles are blazars. Due to Doppler boosting, blazars exhibit extreme stochastic variability. While the origin of this variability is thought to be changes in the accretion flow and jet dynamics, much about blazar variability remains unknown. In this paper we use several blazar-dominated AGN samples to form a catalog of 7918 blazars and candidates – the largest to date. We also collected source types, redshifts, peak frequencies of the spectral energy distribution, radio variability Doppler factors, and X-ray flux densities for as many sources as possible. We used all-sky surveys (CRTS, ATLAS, and ZTF, abbreviated as “CAZ”) to extract their optical multiband flux density on a nightly basis between 2007 and 2023, and we constructed as long and as high cadence light curves as possible for as many sources as attainable. We quantified the variability of the light curves and applied the Bayesian blocks algorithm to determine their flaring periods. The CAZ catalog and light curves as well as the corresponding Bayesian blocks and flaring periods are all provided in the accompanying electronic tables, with the goal of enabling analyses involving jetted AGN variability with unprecedented sample sizes. Overall, we find (1) optical flares generally have a faster rise than decay; (2) optical brightness and variability are strongly dependent on the synchrotron peak frequency; (3) flat spectrum radio quasars and BL Lac objects have comparable optical variability and flare characteristics at the same synchrotron peak frequency; and (4) optical flare times tend to decrease while amplitudes increase with an increasing radio variability Doppler factor.

**Key words.** catalogs – galaxies: active – BL Lacertae objects: general – galaxies: jets – quasars: general – galaxies: statistics

## 1. Introduction

Active galactic nuclei (AGN) exhibit significant non-stellar emission powered by supermassive black holes (SMBH) in the center of their host galaxy. Around a tenth of AGN possess highly collimated jets of relativistic plasma. Blazars constitute a small fraction of jetted AGN, whose jet points toward Earth at small viewing angles ( $\lesssim 10^\circ$ ). This leads to a relativistically boosted broadband nonthermal emission from radio to very-high-energy (0.1–100 TeV)  $\gamma$ -rays, making blazars some of the brightest extragalactic objects in the Universe. For recent reviews on blazars, see Blandford et al. (2019), Böttcher (2019), and Hovatta & Lindfors (2019).

The energy output of blazars in terms of frequency is often studied via their spectral energy distribution (SED). All blazar SEDs show a double-hump structure, the first of which is due to synchrotron radiation of relativistic electrons (typically peaking between the radio and ultraviolet bands). The second hump typically peaks at the X-ray or higher energy bands, and its origin is not fully understood (generally dominated by leptonic inverse

Compton scattering, but hadronic contributions cannot be ruled out; e.g., Marshall et al. 2024; Agudo et al. 2025; Liodakis et al. 2025). Blazars are often categorized based on their synchrotron peak frequency into low-, intermediate-, and high-synchrotron peaked sources, respectively LSP ( $\nu_{\text{sy}} < 10^{14}$  Hz), ISP ( $10^{14} \leq \nu_{\text{sy}} < 10^{15}$  Hz), and HSP ( $\nu_{\text{sy}} \geq 10^{15}$  Hz; e.g., Abdo et al. 2010a).

A more traditional classification of blazars is based on observational features. Flat spectrum radio quasars (FSRQs) are those whose optical spectra exhibit broad emission lines – rest-frame equivalent width (EW) of  $>5 \text{ \AA}$  – and are typically LSPs. On the other hand, BL Lac objects (BLLs) display mostly featureless spectra (EW  $< 5 \text{ \AA}$ ), and their synchrotron emission can peak in a wide range of energy bands, making them LSPs, ISPs, HSPs, or even extreme HSPs (EHSPs; where  $\nu_{\text{sy}} \geq 10^{17}$  Hz; e.g., Costamante et al. 2001). The FSRQs and BLLs are thought to constitute intrinsically different populations, at least partially (e.g., Ghisellini et al. 2010). For example, FSRQs are known to have a broad-line region in the vicinity of their central engine where fast-moving gas clouds absorb and rescatter the light from the accretion disk. On the other hand, the central engine of BLLs is thought to be “naked”. As a result, seed photons for inverse

\* Corresponding author: pouya.kouch@utu.fi

Compton scattering are thought to come from the dusty torus, the broad-line region, or the accretion disk in FSRQs and from synchrotron photons in BLLs. Furthermore, FSRQs and BLLs show different SED behaviors as luminosity increases. In FSRQs, the SED peak frequencies remain constant, with the higher-energy component becoming brighter, whereas in BLLs both SED peaks shift to higher energies (e.g., Ghisellini et al. 2017).

The energy output of blazars, and more generally AGN, exhibits extreme stochastic variability over time (for a recent review, see, e.g., Kankkunen et al. 2025a). As such, their light curves show periods of quiescent and enhanced activity (i.e., “flaring”), during which the energy output can be tens of times greater than the quiescent period. Such flaring periods occur in all bands of the electromagnetic spectrum, although their typical characteristics vary greatly from band to band. For instance, in the radio band, they can be as long as years (e.g., Hovatta et al. 2007; Kankkunen et al. 2025b) but as fast as minutes in higher energy bands, such as  $\gamma$ -rays (e.g., Aharonian et al. 2007; Albert et al. 2007; Pandey & Stalin 2022). In the optical band, blazars show a wide range of timescales, from years to minutes (e.g., Romero et al. 2002; Ruan et al. 2012; Weaver et al. 2020; Jorstad et al. 2022). In general, multiwavelength blazar light curves can be used to investigate the overall jet activity over time (e.g., Lähteenmäki & Valtaoja 2003; León-Tavares et al. 2011), making them a powerful tool for answering some of the biggest open questions about blazars. For example, they are necessary for understanding the origin of the blazar high-energy emission (e.g., Middei et al. 2023; Kouch et al. 2025b) and constraining the different acceleration mechanisms in the jet (e.g., Lioudakis et al. 2022b; Kouch et al. 2024b); for more examples, see the end of the paper (Sect. 6).

In this work, we compile a sample of 7918 AGN (mostly blazars; see Sect. 2) and obtain long-term (as long as  $\sim 15$  yr) and high-cadence (as dense as nightly) flux density light curves in the optical band for as many of them as possible. The light curves are mainly extracted from three all-sky surveys (Sect. 3): the Catalina Real-time Transient Survey (CRTS), the Asteroid Terrestrial-impact Last Alert System (ATLAS), and the Zwicky Transient Facility (ZTF) survey. Together, these surveys form the acronym “CAZ” (CRTS+ATLAS+ZTF), which we use to refer to our catalog of 7918 blazar-dominated AGN and their light curves. This catalog constitutes the most extensive and uniformly compiled sample of blazar optical light curves to date, making it ideal for use in population-based analyses. Nevertheless, we caution that some of the light curves suffer from various complications arising due to automatic forced-photometry (e.g., underestimated errors, presence of extreme outliers, and contamination by their host galaxy and bright nearby sources). In this work we also characterize the variability of the CAZ light curves and identify their flaring periods (Sect. 4), whose results are discussed and interpreted in Sect. 5. Our conclusions are summarized in Sect. 6. Our primary motivation to construct this catalog and its light curves has been to perform the largest-ever population-based spatio-temporal correlation between the blazar and high-energy neutrino populations (e.g., Hovatta et al. 2021; Kouch et al. 2024a, 2025a), which we do in our companion paper, Kouch et al. (2026).

## 2. Compiling the CAZ catalog

We obtain long-term optical light curves for as many blazars and candidates as possible by compiling sources from blazar-dominated catalogs. We begin with the Very Long

Baseline Interferometry (VLBI) based Radio Fundamental Catalog (RFC), which contains  $\sim 22k$  precisely localized compact radio sources. We use the `rfc_2023d`<sup>1</sup> version released in January 2024. Due to the efforts of the VCS (Very Long Baseline Array calibrator surveys; e.g., Kovalev et al. 2007; Gordon et al. 2016) and LBA (Australian Long Baseline Array; e.g., Petrov et al. 2019) observations, the subset of 3236 RFC sources fulfilling  $S_{X\text{-band}}^{\text{VLBI}} \geq 150$  mJy can be considered statistically complete (i.e., this subset contains all compact sources detectable above 150 mJy in the X-band, 8–12 GHz). Since most sources in any flux-limited VLBI sample are expected to be blazars (e.g., Readhead et al. 1978), we begin compiling the CAZ catalog using this subset of 3236 sources. We use this flux-limited subsample over the entire RFC catalog because: (1) it mostly contains confirmed blazars while the entire RFC catalog mostly contains sources of uncertain type, likely several thousand non-blazars; (2) using the entire RFC catalog would not benefit us when constructing the CAZ optical light curves since most of its sources would be too dim in the optical band; and (3) focusing on the flux-limited subsample enables population-based analyses on a statistically complete sample of radio-bright AGN.

Subsequently, we use the latest release of the Fourth Catalog of Active Galactic Nuclei present in the Large Array Telescope (4LAC), 4LAC-DR3<sup>2</sup>, which was released in August 2022 and is based on 12 years of data from the *Fermi* Large Array Telescope (LAT; Atwood et al. 2009). We add all 3814 low- and high-latitude 4LAC sources into the CAZ source list, using their radio counterpart coordinates provided in the 4LAC catalog (Ajello et al. 2022). Additionally, as seen below, we tabulate several other parameters (e.g., redshift, SED peaks) from the 4LAC catalog. We then use the third catalog of HSP blazars (3HSP<sup>3</sup>), last updated in April 2020. We add all its 2013 sources into the CAZ source list, while adopting the counterpart coordinates. As discussed in Chang et al. (2019), most of the counterparts in the 3HSP catalog are from the WISE<sup>4</sup> all-sky survey (Wide-field Infrared Survey Explorer at IPAC); however, some only have 5BZC<sup>5</sup> (the fifth Roma-BZCAT Multifrequency Catalogue of Blazars) or *Fermi*-LAT (3LAC) counterparts. We tabulate the synchrotron peak frequency and redshift of the sources whenever provided by the 3HSP catalog. Next, we add all 1625 sources in the Candidate Gamma-Ray Blazar Survey (CGRaBS<sup>6</sup>; Healey et al. 2008) source catalog, which is complete down to 65 mJy flux density at 4.8 GHz and radio spectral index  $\alpha > -0.5$  where  $S \propto \nu^\alpha$  (Richards et al. 2011). We adopt the redshift and source classification of this catalog. Finally, we add all 3561 sources of the aforementioned 5BZC catalog of blazars, while tabulating their redshift and 5BZC classification (Massaro et al. 2015).

In the next step, we cross-check the radio-counterpart RA and Dec. coordinates of the sources to the closest  $0.001^\circ$  (comparable to the typical seeing expected from the all-sky surveys), leaving us with 8003 unique coordinates. We further identify duplicate sources by, first, using the available data in the cross-matched catalogs (e.g., looking for identical radio-counterparts), and then manually inspecting the known properties (e.g., redshift, source type) of all 140 pairs of sources closer than  $0.2^\circ$ . Sources suspected to be duplicates are then merged, with their

<sup>1</sup> [https://astrogeo.org/sol/rfc/rfc\\_2023d/](https://astrogeo.org/sol/rfc/rfc_2023d/)

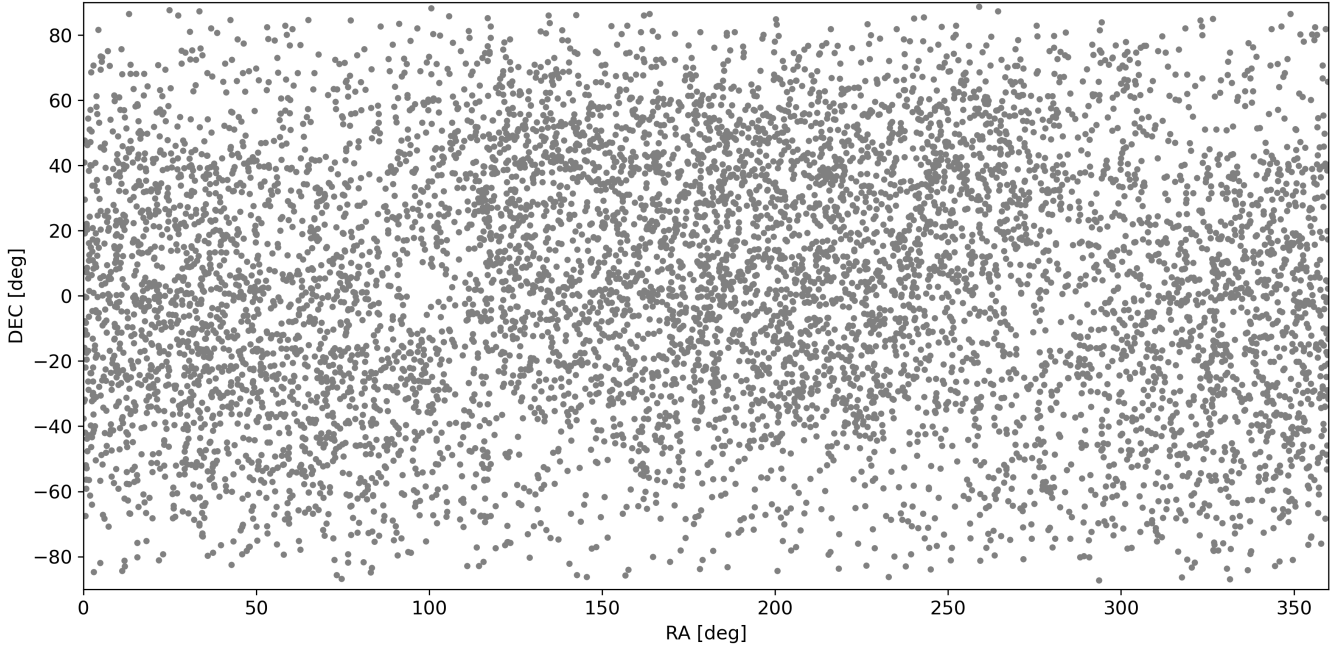
<sup>2</sup> <https://fermi.gsfc.nasa.gov/ssc/data/access/lat/4LACDR3/>

<sup>3</sup> <https://www.ssdsc.asi.it/3hsp/>

<sup>4</sup> <https://wise2.ipac.caltech.edu/docs/release/allsky/>

<sup>5</sup> <https://www.ssdsc.asi.it/bzcat/>

<sup>6</sup> <https://heasarc.gsfc.nasa.gov/W3Browse/radio-catalog/cgrabs.html>



**Fig. 1.** Sky distribution of 7918 sources in the CAZ catalog using the equatorial coordinate system (J2000 epoch).

final RA and Dec. preferentially corresponding to their radio counterpart from: RFC → 4LAC → 3HSP → CGRaBS → 5BZC. We note that, as seen below, we remove two sources (CGRaBS J1537–7154 and RFC J1911+0458) from the catalog for not being AGN.

The final CAZ catalog contains 7918 unique AGN and candidates, of which 5111 (64.5%) are in the RFC catalog, 3225 (40.7%) are in the statistically complete, flux-limited RFC subsample<sup>7</sup> referred to as RFC<sup>†</sup>, 3814 (48.2%) are in the 4LAC-DR3 catalog, 2013 (25.4%) are in the 3HSP catalog, 1625 (20.5%) are in the CGRaBS sample, and 3561 (45.0%) are in the 5BZC catalog. Moreover, at least 696 (8.8%) of these 7918 sources have been a monitoring target in one of the major blazar monitoring programs<sup>8</sup>. The sky locations of the sources in the CAZ catalog are shown in Fig. 1.

For as many of the 7918 sources in the CAZ catalog as possible, we collect source types, redshifts, SED peak frequencies, and radio variability Doppler factors mainly from the above catalogs. We also extract X-ray flux densities when possible. Additionally, we extract redshift and source type information from the Set of Identifications, Measurements, and Bibliography for Astronomical Data (SIMBAD<sup>9</sup>) for as many sources as possible.

The above catalogs and SIMBAD have somewhat different approaches of classifying their sources. Therefore, we devised our own AGN classes to unify the distinct classifications. Our blazar-focused AGN classification scheme is shown below:

- “Q”: FSRQs and FSRQ candidates
- “B”: BLLs and BLL candidates
- “G”: Host-galaxy-dominated BLLs<sup>10</sup>

<sup>7</sup> The initial RFC flux-limited subsample has 3236 sources meeting  $S_{X\text{-band}}^{\text{VLBI}} \geq 150$  mJy; however, ten of them have neighboring sources closer than  $0.001^\circ$  and one of them is not an AGN. These 11 are removed from the CAZ catalog.

<sup>8</sup> <https://www.cv.nrao.edu/MOJAVE/blazarlist.html>

<sup>9</sup> <https://simbad.cds.unistra.fr/simbad/>

<sup>10</sup> This is a 5BZC exclusive category which we do not merge with any other blazar category.

- “U”: Blazars of unknown type and blazar candidates
- “A”: Confirmed non-blazar AGN<sup>11</sup> and AGN candidates

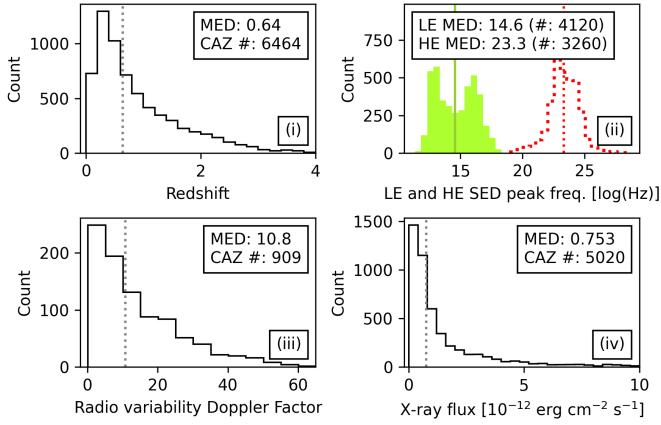
Many sources have classification information from multiple references. We accounted for this by prioritizing the catalogs as 4LAC → CGRaBS → 3HSP → 5BZC → SIMBAD. We note that the 3HSP catalog does not have source classification, but all of its sources are identified as HSPs, making them BLLs. If a radio source does not have a classification from any of the catalogs, it is taken as an AGN candidate. The 7918 AGN and candidates in the CAZ catalog fit into our blazar-focused classification as shown below:

- Q: 2495 (31.5%)
- B: 2726 (34.4%)
- G: 112 (1.4%)
- U: 1697 (21.4%)
- A: 888 (11.2%)

Therefore, 88.8% of the 7918 sources in the CAZ catalog are blazars or blazar candidates. We included the 11.2% non-blazar AGN and AGN candidates because the classifications from different catalogs do not always agree, so it is possible that some of the currently classified non-blazar AGN are actually blazars. Furthermore, by minimally excluding sources, we maintain continuity between the CAZ and its parent catalogs (i.e., 4LAC, 3HSP, CGRaBS, 5BZC, and the statistically complete subsample of the RFC), making future comparative studies between them more straightforward.

There are a total of 6464 redshift estimates for the 7918 sources (i.e., 82.7% of the CAZ catalog). These are primarily collected from the catalogs mentioned above as well as the Optical Characteristics of Astrometric Radio Sources (OCARS) catalog, which contains both spectroscopic and photometric redshift estimates (Malkin 2018). Lastly, we also search SIMBAD for redshift estimates. Since the reliability of these estimates differ and there are often mismatching estimates, we select the final redshift for each source using the following priority order: 4LAC → CGRaBS → 5BZC → SIMBAD → OCARS → 3HSP. This

<sup>11</sup> This includes radio galaxies, Seyfert type galaxies, etc.



**Fig. 2.** Distributions of (i) redshift, (ii) LE and HE peak frequencies (shown in green and red, respectively), (iii) radio variability Doppler factor, and (iv) median X-ray flux density of sources in the CAZ catalog. “MED” refers to median, and “CAZ #” refers to the number of available parameters.

priority order is identical to the classification priority order with the exception of 3HSP. The 4LAC catalog has the highest priority because all of its redshift estimates are spectroscopic and their reliability has been checked rigorously, while the 3HSP catalog has the lowest priority because it primarily reports photometric redshifts. The distribution of the 6464 redshift estimates is shown in Fig. 2i.

We have a total of 4120 (52.0% of 7918) low-energy (LE; i.e., synchrotron) peak frequencies from the 4LAC (2825 of 4120) and 3HSP (1295 of 4120) catalogs, where the former value is prioritized in case of duplicates. In addition, there are a total of 3260 (41.2% of 7918) high-energy (HE) peak frequencies from the 4LAC catalog. The distributions of these LE and HE peak frequencies are shown in Fig. 2ii. Based on these LE peak frequencies, we consequently identify 1699 sources (41.2% of 4120) as LSPs, 536 (13.0%) as ISPs, 1619 (39.3%) as HSPs, and 266 (6.5%) as EHSPs using the synchrotron peak frequency thresholds defined in Sect. 1.

Moreover, we collected 909 (11.5% of 7918) radio variability Doppler factors for the sources in the CAZ catalog. The majority of these (875 of 909) are from Liodakis et al. (2018a), and the remaining 34 are calculated in this study using previously unavailable redshift information via Eq. (2) of Liodakis et al. (2018a). The distribution of these radio variability Doppler factors is shown in Fig. 2iii.

Lastly, we obtained a median X-ray flux density for 5020 (63.4% of 7918) sources in the CAZ catalog. To do this, in our companion paper Paggi et al. (in prep.), we collect X-ray light curves for as many of the CAZ sources as possible using archival data from the space-borne *Neil Gehrels Swift* Observatory (*Swift*-XRT<sup>12</sup>; Gehrels et al. 2004), *Chandra* X-ray Observatory<sup>13</sup> (e.g., Weisskopf et al. 2002), and X-ray Multi-Mirror Mission (*XMM-Newton*<sup>14</sup>; Jansen et al. 2001). We then calculated the median of the detected ( $3\sigma$ ) X-ray flux densities in these light curves. For 5020 sources in the CAZ catalog, the X-ray light curve has at least one detection. This leads to 5020 median flux density measurements, whose distribution is shown in Fig. 2iv. For 734 (9.3% of 7918) sources in the CAZ catalog, we only find upper limits on the X-ray flux density. More details

about the X-ray data extraction is given in Paggi et al. (in prep.), where we also provide the X-ray light curves and spectra of these sources. In that paper, we also construct an X-ray-selected, statistically complete sample of blazars and perform a number of complementary blazar-population analyses.

### 3. CAZ light curves

Here we describe the process of obtaining optical light curves for the 7918 sources in the blazar-dominated CAZ catalog. In short, we constructed decade-long and daily cadenced optical light curves for as many sources as possible, mainly using three all-sky surveys: CRTS, ATLAS, and ZTF (Sect. 1). We additionally obtained 408 white dwarf light curves (referred to as “WD” CAZ light curves), which are used as non-variable control light curves in the ensuing variability analysis (Sect. 4). The CAZ light curves have data until the end of 2023, i.e., Modified Julian Date (MJD) of 60310. The final CAZ light curves are provided as electronic tables.

#### 3.1. Extracting raw data

We collected data from the CRTS<sup>15</sup> all-sky survey, which officially ran from 2007 to 2016 (Drake et al. 2009). This survey is obtained without a filter (i.e., closest to the standard V filter with  $\lambda_{\text{eff}} = 551$  nm) via three telescopes: two located in Arizona, with limiting magnitudes of 19.5 and 21.5, and one in Australia with a limiting magnitude of 19.0. Subsequently, we collected data from the ATLAS<sup>16</sup> all-sky survey, running since 2016 (Tonry et al. 2018). It primarily observes in filters cyan (c;  $\lambda_{\text{eff}} = 533$  nm) and orange (o;  $\lambda_{\text{eff}} = 578$  nm) via four telescopes: two located in Hawaii, one in Chile, and one in South Africa. It also occasionally observes in g ( $\lambda_{\text{eff}} = 475$  nm), r ( $\lambda_{\text{eff}} = 637$  nm), and i ( $\lambda_{\text{eff}} = 783$  nm) filters. The typical limiting magnitude of ATLAS has been  $\sim 19$  since 2019 (prior to 2019 it was dimmer; Smith et al. 2020). We then collected data from the ZTF<sup>17</sup> all-sky survey, running since 2018 (Bellm et al. 2019). We used its 20<sup>th</sup> data release, containing data between March 2018 and Jan 2024. ZTF observes using the Palomar 48” Schmidt telescope located in California in three filters: g, r, and i. The limiting magnitude for g and r is  $\sim 21.5$ , and for i  $\sim 21.0$ . In addition to these all-sky survey light curves, we included data from our Tuorla blazar monitoring program<sup>18</sup> and the Katzman Automatic Imaging Telescope (KAIT<sup>19</sup>) monitoring program. As part of the Tuorla program we have been monitoring 205 blazars in the R-band ( $\lambda_{\text{eff}} = 658$  nm) since 2002 via various telescopes (Nilsson et al. 2018), and in this paper we publish our post-2013 data for the first time. KAIT has been searching for supernovae in the V filter since 1997 (Filippenko et al. 2001) and has ended up monitoring 155 blazars serendipitously (e.g., Cohen et al. 2014; Liodakis et al. 2018b).

While for the two dedicated monitoring programs, source-based light curves in flux density versus MJD are available, in the case of the three all-sky surveys forced photometry at each source coordinate (within  $\pm 0.001^\circ$ ) is needed to obtain the light curves. If only magnitudes are available, we convert them to flux densities using  $S = S_0 \cdot 10^{-0.4m}$  where  $S$ ,  $S_0$ , and  $m$  are flux

<sup>12</sup> <https://heasarc.gsfc.nasa.gov/>

<sup>13</sup> <http://cda.harvard.edu/chaser>

<sup>14</sup> <http://nxsa.esac.esa.int/nxsa-web>

<sup>15</sup> <http://crts.caltech.edu/>

<sup>16</sup> <https://fallingstar-data.com/>

<sup>17</sup> <https://ztf.caltech.edu/>

<sup>18</sup> <https://tuorlablazar.utu.fi/>

<sup>19</sup> <https://w.astro.berkeley.edu/bait/kait.html>

density, flux density zero-point, and apparent magnitude, respectively. We use the following flux density zero-points for each band:  $S_{0,V} = 3640$  Jy,  $S_{0,R} = 3080$  Jy (Mead et al. 1990), and  $S_{0,g,r,i} = 3631$  Jy (referring to the monochromatically defined  $g$ ,  $r$ , and  $i$  AB-magnitudes; Tonry et al. 2012).

The light curve quality across these all-sky surveys and dedicated monitoring programs differs greatly. While the latter are generally of higher quality due to supervised data collection and reduction, the all-sky surveys (especially ATLAS) suffer from outliers and low signal-to-noise ratio (S/N) observations due to automatic forced-photometry and data reduction algorithms. Manual inspection of the raw light curves reveals that ATLAS light curves exhibit unrealistically bright data points (e.g., the brightest ATLAS data point is around  $-38$  magnitudes). Since the brightest raw data points from all other surveys and programs are  $\gg 5$  magnitudes, we exclude all ATLAS raw data points brighter than 5 magnitudes, corresponding to  $\sim 0.6\%$  of  $\sim 18.9$ M ATLAS data points. Subsequently, we impose  $S/N > 1$  on all extracted light curves, which excludes  $\sim 0.01\%$  of  $\sim 1.6$ M CRTS raw data points,  $\sim 11.7\%$  of  $\sim 18.9$ M ATLAS,  $\sim 0.01\%$  of  $\sim 6.3$ M ZTF,  $\sim 0.01\%$  of  $\sim 37$ k KAIT, and none of  $\sim 45$ k Tuorla data points. We note that these estimates include the 408 WD light curves in addition to the 7918 AGN light curves. More strict detection criteria ( $S/N \geq 2$ ) are imposed after cleaning and nightly binning the light curves (see Sect. 3.2).

The majority of the CRTS raw data is taken from the online repository<sup>20</sup>. In addition to this, we have access to extended CRTS light curves (by  $\sim 1$  yr for some sources; as early as mid-2005), which we add onto their public data. The ATLAS<sup>21</sup> and ZTF<sup>22</sup> raw data are also extracted using their respective public repositories. Overall, 5185 (65.5% of 7918) sources have at least some data points in their extracted CRTS light curves, all 7918 in their ATLAS light curves, and 5607 (70.8% of 7918) in their ZTF light curves. From the Tuorla sample, there are 205 blazar light curves present, while from the KAIT sample there are 155.

### 3.2. Cleaning outliers and nightly binning

While blazars are highly variable, even on timescales of minutes, the above surveys and monitoring programs cannot detect such fast, intranight variability reliably. Therefore, we assume that all intranight variability is due to noise and bin all the flux densities on a nightly basis. Before binning, we first ensure that a filter has at least three data points across one source, then we use the criteria introduced by Hovatta et al. (2014) to identify and remove outliers: (1) we find the mean and standard deviation of the intranight magnitude values, (2) if standard deviation is greater than 0.5 magnitudes, we exclude the entire night, otherwise we select the brightest and the dimmest data points, (3) if either the brightest or dimmest data is different from the mean by more than 0.8 magnitudes, we exclude the more extreme of the two, and (4) we repeat the process until there are no intranight magnitude measurements with a standard deviation  $> 0.5$ .

Subsequently, we calculated a binned flux density for each night (with  $S_i \pm \sigma_i$  flux densities and errors, where  $i = 1, \dots, N$ ) using inverse-variance weighted averaging:

$$\bar{S} = \frac{\sum_{i=1}^N S_i w_i}{\sum_{i=1}^N w_i} \quad (1)$$

<sup>20</sup> [http://nunuku.caltech.edu/cgi-bin/getmulticonedb\\_release2.cgi](http://nunuku.caltech.edu/cgi-bin/getmulticonedb_release2.cgi)

<sup>21</sup> <https://fallingstar-data.com/forcedphot/apiguide/>

<sup>22</sup> [https://irsa.ipac.caltech.edu/docs/program\\_interface/ztf\\_api.html](https://irsa.ipac.caltech.edu/docs/program_interface/ztf_api.html)

$$\sigma_{\bar{S}} = \sqrt{\frac{1}{\sum_{i=1}^N w_i}}, \quad (2)$$

where  $w_i = \sigma_i^{-2}$  and  $\bar{S}$  is the binned flux density, with  $\sigma_{\bar{S}}$  being its error. To ensure the robustness of the nightly binned errors ( $\sigma_{\text{night}}$ ), we added 1.2% of  $\bar{S}$  to  $\sigma_{\bar{S}}$  in quadrature as a flat-fielding factor via  $\sigma_{\text{night}} = [\sigma_{\bar{S}}^2 + (0.012 \times \bar{S})^2]^{1/2}$  (e.g., Nilsson et al. 2018). We excluded all nights with  $S/N = \bar{S}/\sigma_{\text{night}} < 2$ . If there was only one data point per night, we required  $S/N \geq 3$ . The above criteria exclude  $\sim 3.0\%$  of  $\sim 1.6$ M CRTS data points (after implementing the basic exclusion criteria of Sect. 3.1),  $\sim 7.8\%$  of  $\sim 16.6$ M ATLAS,  $\sim 2.4\%$  of 6.3M ZTF,  $\sim 11.8\%$  of  $\sim 37$ k KAIT, and  $\sim 2.7\%$  of  $\sim 45$ k Tuorla.

### 3.3. Merging all filters

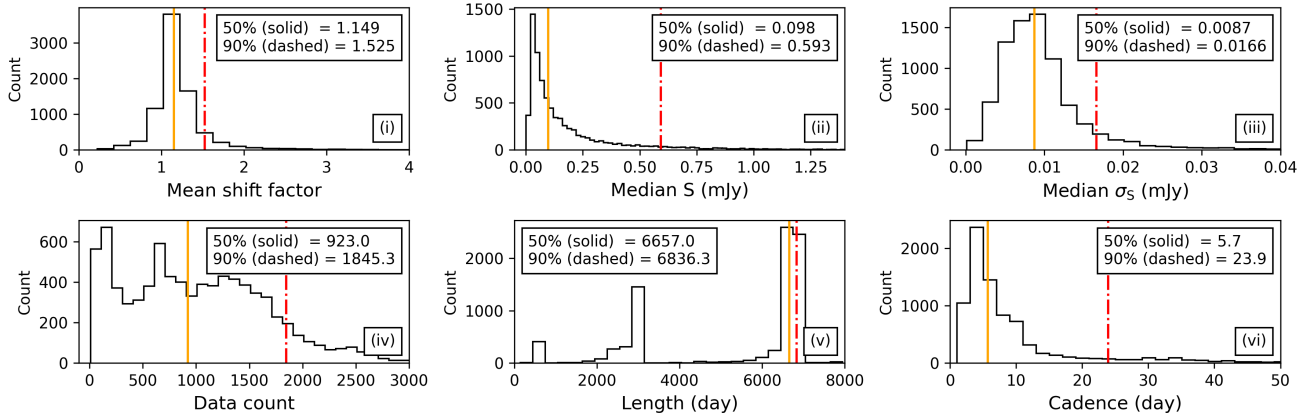
All the above treatments were applied to data points within one filter of a given survey or monitoring programs. There are a total of 11 unique survey and filter combinations: one from CRTS, denoted as: [C]V; one from KAIT, [K]V; one from Tuorla, [T]R; five from ATLAS, [A]o, [A]c, [A]r, [A]g, and [A]i; and three from ZTF, [Z]r, [Z]g, and [Z]i. We merge these in order to maximize the cadence of the final CAZ light curves. Merging is ideally done via empirically well-established filter transformation relations when the spectral shape of a source is known. Unfortunately, we neither have the filter transformation relations, nor the source-specific spectral shapes. Another possibility is to use simultaneous multicolor data to calculate a filter transformation relation<sup>23</sup>. However, such relations must be found uniquely for each source, because blazar emission can be strongly chromatic with a wide range of spectral shapes from one source to another and show varying degrees of chromaticity at different brightness levels (i.e., bluer-when-brighter and redder-when-brighter trends; e.g., Rani et al. 2010; Bonning et al. 2012; Sandrinelli et al. 2014; Zhang et al. 2015; Meng et al. 2018; Otero-Santos et al. 2022). Unfortunately, due to the lack of simultaneous multicolor data, we cannot do this either. Thus, we choose not to perform filter transformations but translate the filters in flux density space by multiplying their flux densities and errors<sup>24</sup> with a shift factor, determined automatically on a filter-by-filter and source-by-source basis. This approach reliably eliminates overall flux density discrepancies arising due to a constant degree of chromaticity (e.g., when a filter is brighter than another by a constant factor on average). However, it cannot reliably account for flux density discrepancies which arise when the degree of chromaticity changes over time (e.g., due to bluer-when-brighter or redder-when-brighter trends). Nevertheless, translating the filters seems to preserve the variability information of most CAZ light curves rather well, making it suitable for our purposes.

The shift factor for a filter and survey pair (two of the aforementioned 11, e.g., [S<sub>1</sub>]f<sub>i</sub> and [S<sub>2</sub>]f<sub>j</sub>; hereafter, simply referred to as a filter) was determined via one of four techniques, in decreasing order of priority:

- the average ratio of at least five simultaneous (within  $\pm 1$  d) flux densities,

<sup>23</sup> For example, to convert  $g$  and  $r$  AB-magnitudes to Cousins V-band magnitude, the linear relationships between  $m_g - m_r$  versus  $g$  and  $m_g - m_r$  versus  $r$  are needed, respectively (e.g., Tonry et al. 2012).

<sup>24</sup> Multiplying the error bar ensures that the S/N of the translated data point remains constant.



**Fig. 3.** Global distributions of 8130 merged CAZ light curves (7722 of AGN and 408 of WDs) with  $\geq 10$  data points and  $\geq 30$  d duration. The panels are as follows: (i) mean shift factor across all filters per light curve. (ii) Median flux density per light curve. (iii) Median flux density error per light curve. (iv) Data count per light curve. (v) Length of each light curve. (vi) Cadence per light curve calculated by dividing the length of each light curve by its data count. The vertical solid orange line shows the median, and the dashed red line shows the 90th percentile.

- same as above, except  $\pm 2$  d is used to define simultaneous flux densities,
- the ratio of the average flux density of one filter to the other within their overlapping time window, or
- same as above, except the overlapping time window is calculated with a  $\pm 50$  d extension.

In the last two techniques, to calculate the overlapping flux density averages, there must exist at least five overlapping data points for each filter. Otherwise, they are not shifted at all. We began by listing all possible filter pairs for a light curve. We then checked if any of them can be shifted using the highest priority technique. If yes, one is shifted onto the other, and they are merged into one:  $[S_1]f_i + [S_2]f_j$ . A merged filter was then treated as an independent filter and sought to be merged with other filters. If a shift technique does not result in any mergers, we applied the next shift technique in the priority list. This is repeated until either all filters in the light curve were merged into one or when the last shift technique cannot result in any mergers. Globally, 86.4, 7.2, 5.5, and 0.9% of the shifts are done using the four techniques in decreasing order of priority, respectively. Whenever a filter cannot be shifted, we reported its original flux densities. Finally, the reference filter (i.e., the filter onto which all other filters are shifted) for each light curve is chosen based on data availability. In decreasing order of priority, we chose the reference filter to be [Z]r, [Z]g, [Z]i, [C]V, [T]R, [K]V, [A]o, [A]c, [A]r, [A]g, and [A]i. We report the final shift factor for each filter in the electronic table. A shift factor of 1 implies that the filter is not shifted (e.g., the reference filter). Notably, all shifted filters can be shifted back to their original flux density level by dividing their flux densities and flux density errors with the final shift factor we provide in the electronic table.

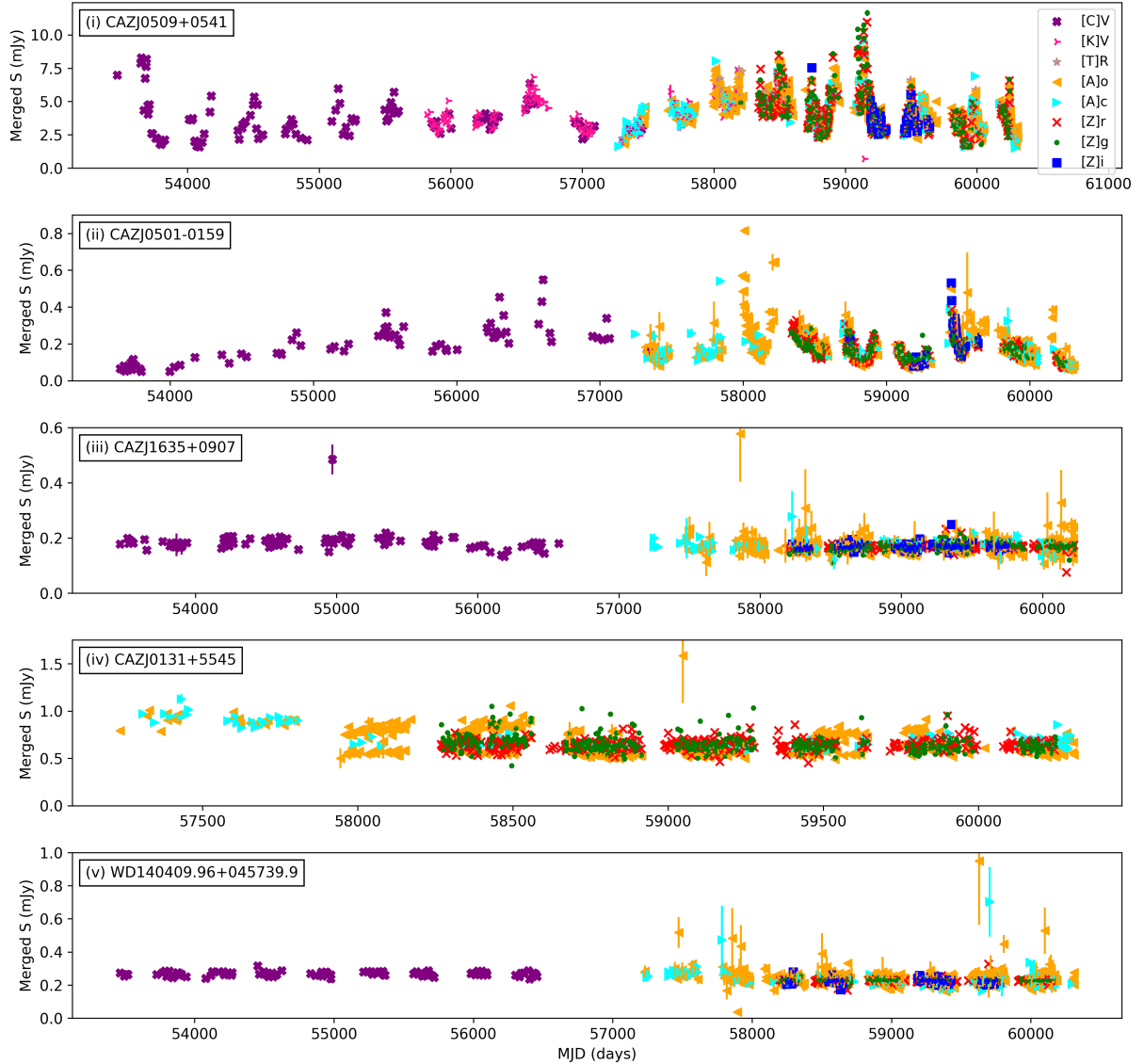
In Fig. 3, we show the global distribution of average shift factor across filters, median flux density, median flux density error, data count, length, and cadence per light curve for the merged light curves with  $\geq 10$  data points and  $\geq 30$  d duration (corresponding to 7722 sources in the CAZ catalog and 408 WDs). The merged light curves typically have a total of  $\sim 900$  data points spanning  $\sim 18$  years in length, translating to an average weekly cadence. However, we emphasize that deviations from these typical values often happen, and these global averages do not visualize potential source-by-source caveats (e.g., large seasonal gaps, noise-like behavior, outliers). Notably, combining three all-sky surveys creates three distinct global peaks in the distributions of the data count and length.

In Fig. 4 we show five example merged light curves: from top to bottom, the light curves of CAZJ0509+0541 (TXS 0506+056; a highly variable blazar), CAZJ0501–0159 (another variable blazar), CAZJ1635+0907 (a non-variable blazar), CAZJ0131+5545 (a host-galaxy-dominated blazar), and WD140409.96+045739.9 (a white dwarf) are plotted. In Fig. 5 we present zoomed-in versions of these light curves where we also show a few of their identified periods of enhanced emission and flares (see Sect. 4). As evident, some of the light curves are of much lower quality than others. While the merging process can introduce some artifacts, the most notable problems stem from the raw data. For example, as seen at around MJD 58 000 in panel iv of Fig. 4 (also, see panel vii of Fig. 5), the light curve of the host-galaxy-dominated source CAZJ0131+5545 shows two states of constant flux. This is due to the raw ATLAS light curves, which often show such problematic behavior around that epoch in host-galaxy-dominated light curves. While we do not know their cause, we attempt to mitigate their effect on the results of this study at various steps (see, e.g., Sect. 4.1). In a future iteration of this work, we aim to use specialized tools designed to mitigate issues with ATLAS forced photometry, such as ATClean (Rest et al. 2023), of which we were not aware at the time of constructing the CAZ light curves of this work.

#### 4. Characterizing variability and flaring behavior

In this section we describe how we quantify the overall variability of the merged CAZ light curves using fractional variability (e.g., Sokolovsky et al. 2017; see Sect. 4.1), and how we identify periods of enhanced emission and flaring periods in them utilizing the Bayesian block method (Scargle et al. 2013; see Sects. 4.2 and 4.3). We provide the fractional variability measures and flaring periods of the CAZ light curves in the electronic tables.

Since complex color relations and differences in observational properties of the filters (e.g., survey sensitivities, aperture sizes, typical error sizes, cadence) introduce variability upon merging, we control for such non-intrinsic variability using WDs. While the flux density of WDs can exhibit rather small intrinsic variability (on the order of a few percent of their average flux density; e.g., Bell 2026), they are non-variable in comparison to blazars.



**Fig. 4.** Example merged light curves. The name of the source is given on the top-left corner of each panel. We note that some light curves are of much higher quality than others (e.g., plot i compared to iv; the former is of the well-known, highly variable TXS 0506+056, while the latter is of a host-galaxy-dominated source). A zoomed-in version of these light curves (including identified periods of enhanced emission and flares) is given in Fig. 5.

#### 4.1. Quantifying variability

Fractional variability ( $F_{\text{var}}$ ) is a commonly used variability measure for time series data, especially for AGN and blazar light curves (e.g., Gliozzi et al. 2003; MAGIC Collaboration 2024). It quantifies the average deviation of the data in excess of the observational errors. For a sequence of  $S_i$  flux densities, where  $i = 1, \dots, N$ , it is defined as

$$F_{\text{var}} = \sqrt{\frac{\left[ \frac{1}{N-1} \sum_{i=1}^N (S_i - \bar{S})^2 \right] - \sigma^2}{\bar{S}^2}}, \quad (3)$$

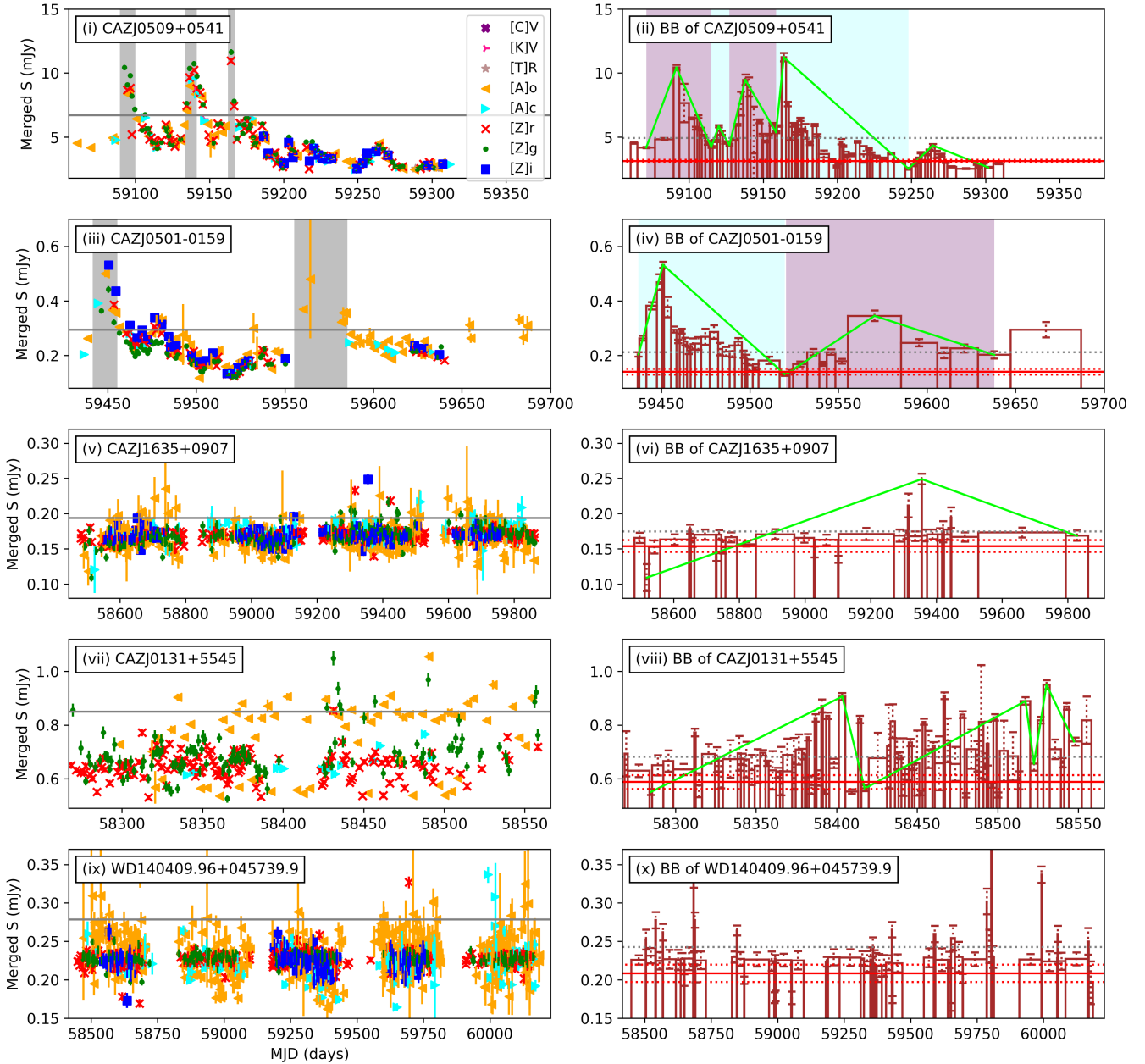
and its error is calculated via

$$\Delta F_{\text{var}} = \sqrt{F_{\text{var}}^2 + \sqrt{\left( \sqrt{\frac{2}{N}} \cdot \frac{\sigma^2}{\bar{S}^2} \right)^2 + \left( \sqrt{\frac{\sigma^2}{N}} \cdot \frac{2F_{\text{var}}}{\bar{S}} \right)^2}} - F_{\text{var}}, \quad (4)$$

where  $\bar{S} = \frac{1}{N} \sum_{i=1}^N S_i$  and  $\sigma^2 = \frac{1}{N} \sum_{i=1}^N \sigma_i^2$  are the sample mean and mean square error, respectively (Vaughan et al. 2003; Poutanen et al. 2008).

In Fig. 6, we show  $F_{\text{var}}$  of the merged CAZ, CRTS-, ATLAS-, and ZTF-only blazar and WD light curves against their median flux densities. The merged WDs appear to show the same variability as the blazar population, which is unexpected. This problematic behavior is also seen when ATLAS-only merged filters are considered. On the other hand, in both CRTS- and ZTF-only plots, the WDs cluster at smaller variability measures in comparison to majority of the blazar population. This suggests that ATLAS light curves are prone to noise and, since they typically have more data points than the other surveys, their inclusion contaminates the merged CAZ light curves.

In addition to  $F_{\text{var}}$ , we explore quantifying the variability of the merged (CAZ, CRTS-, ATLAS-, and ZTF-only) light curves using three other variability measures: interquartile range (IQR; e.g., Sokolovsky et al. 2017), inverse von Neumann (1941) ratio ( $1/\eta$ ; e.g., Shin et al. 2009), and excursions ( $E_x$ ;

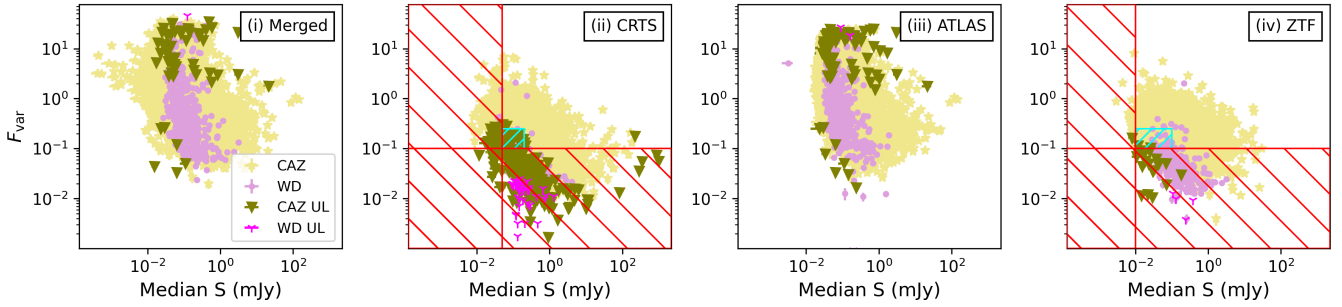


**Fig. 5.** Example zoomed-in light curves with BB95 periods (Sect. 4.2) and BBHOP flares (Sect. 4.3) identified. The merged light curves on the left are zoomed-in versions of those shown in Fig. 4. The light curves on the right are their corresponding BB light curves. The gray vertical area on the left shows BB95 periods, which are BBs whose flux density is larger than  $S_{95\%}$  (shown as the gray horizontal line in the left). BBHOP flares are shown on the right using green lines, connecting their start to the peak and then to the end BB. Prominent BBHOP flares are further distinguished with a purple or cyan background. The horizontal solid red line in the right panels represents the reference flux density for determining the global amplitude  $\mathcal{G}$  (i.e.,  $S_{25\%}$ ; the dotted lines represent its error). The gray horizontal dotted line in the right plots represents  $S_{75\%}$ , which is needed when determining prominent BBHOP flares. We note that CAZJ1635+0907 and CAZJ0131+5545 were determined to be non-variable, and WD140409.96+045739.9 is non-variable by definition. Thus, we did not identify BB95 and prominent BBHOP flares in their light curves.

Sokolovsky et al. 2017). We describe these in more detail and plot their variability estimate against median flux density in Appendix A. While the ATLAS WD scatter is slightly more contained in IQR as compared to  $F_{\text{var}}$ , it appears to show a problematic more-variable-when-brighter trend. With  $1/\eta$ , the ATLAS WD scatter is well contained; however, all surveys exhibit the more-variable-when-brighter behavior. While  $E_x$  does not show such behavior, similar to  $F_{\text{var}}$ , it suffers from a large scatter in the ATLAS WDs. Therefore, as  $F_{\text{var}}$  is a well-established variability measure for blazars (Schleicher et al. 2019), we opt to use it as

the primary metric to quantify the overall variability of the CAZ light curves. However, due to the issues introduced by ATLAS, we only use the  $F_{\text{var}}$  of CRTS- and ZTF-only merged light curves to define global variability criteria.

We followed the methodology of Hovatta et al. (2014) to define non-variability criteria in Eq. (5). CRTS- or ZTF-only merged light curves with a flux density lower than  $S_{\text{low}}$  or  $F_{\text{var}}$  lower than  $F_{\text{var,low}}$  are considered non-variable. Additionally, light curves that fall within a box-like region defined by  $S_{\text{box}}$  and  $F_{\text{var,box}}$  are considered non-variable for being WD-like. In other



**Fig. 6.** Fractional variability ( $F_{\text{var}}$ ) versus median flux density plots. The panels are as follows: (i) Merged light curves. (ii) CRTS only. (iii) ATLAS only. (iv) ZTF only. The term “UL” refers to  $3\sigma$  upper limits of  $F_{\text{var}}$ . The red hatched areas in (ii) and (iv) represent the low flux density and low variability regimes to determine non-variable light curves. The cyan hatched box is typically occupied by WDs and is another non-variable regime.

words, if either of the following conditions holds, the CRTS- and ZTF-only light curves are deemed non-variable:

$$\begin{cases} \tilde{S} < S_{\text{low}} \\ F_{\text{var}} < F_{\text{var,low}} \\ (\tilde{S} < S_{\text{box}}) \text{ AND } (F_{\text{var}} < F_{\text{var,box}}), \end{cases} \quad (5)$$

where  $\tilde{S}$  refers to the median flux density. In the CRTS-only case,  $S_{\text{low}} = 0.05$  mJy,  $F_{\text{var,low}} = 0.10$ ,  $S_{\text{box}} = 0.20$  mJy, and  $F_{\text{var,box}} = 0.25$ . In the ZTF-only case,  $S_{\text{low}} = 0.01$  mJy,  $F_{\text{var,low}} = 0.10$ ,  $S_{\text{box}} = 0.10$  mJy, and  $F_{\text{var,box}} = 0.25$ . These non-variable regimes are optimized by eye and are marked on the CRTS- and ZTF-only plots in Fig. 6.

The criteria of Eq. (5) only determine if the CRTS or ZTF portions of a CAZ light curve are variable or not. To confidently identify a CAZ light curve as variable, we required either CRTS- or ZTF-only portions to be variable with  $F_{\text{var}} > 0.30$ . If  $F_{\text{var}} \leq 0.30$ , then they both need to be variable simultaneously for a confident variable identification. This extra condition ensures that a slight outlier-induced increase in the  $F_{\text{var}}$  of either CRTS- or ZTF-only portions does not falsely identify the source as variable. Notably, without CRTS or ZTF data, we cannot reliably establish the variability of a CAZ light curve. This means that light curves with only ATLAS data are conservatively assumed to be non-variable. When applied to 7722 CAZ light curves with at least some CRTS or ZTF data, which have  $\geq 10$  data points and are  $\geq 30$  d in duration, these criteria identify 2798 (36.2%) as confidently variable.

#### 4.2. Identifying periods of enhanced emission

The Bayesian block algorithm (Scargle et al. 2013) is a widely used technique for characterizing flux variations and identifying flares in blazar light curves (e.g., Wagner et al. 2022; de Jaeger et al. 2023). It is especially powerful for systematically characterizing flares in a large number of light curves, due to its ability to adaptively bin a light curve using only one easily quantifiable Bayesian prior. Therefore, in this subsection we use it on the merged CAZ light curves (see Sect. 3.3) to identify periods of enhanced emission and, in Sect. 4.3, flaring periods. The Bayesian blocks corresponding to the CAZ light curves are provided as electronic tables.

We apply the Bayesian block algorithm using its implementation in the Python package *astropy* (Astropy Collaboration 2022) to 7722 merged CAZ light curves with  $\geq 10$  data points and  $\geq 30$  d duration. The algorithm then adaptively determines instances of significant flux density variation, between which lie periods (i.e., “blocks”) of insignificant variation. We bin the

flux density within each Bayesian block (BB). The binned flux densities are obtained via Eq. (1) and their error via

$$\sigma_{\bar{S},\text{BB}} = \max\left(\tilde{\sigma}_i, \sqrt{\frac{\sum_{i=1}^N (S_i - \bar{S})^2 \cdot w_i}{\sum_{i=1}^N w_i}}\right), \quad (6)$$

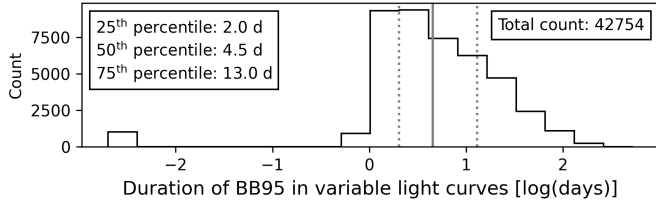
where  $\tilde{\sigma}_i$  is the median flux density error within a BB and, similar to Eq. (1),  $w_i = \sigma_i^{-2}$  and  $\bar{S}$  is the binned flux density for each BB. Equation (6) ensures that  $\sigma_{\bar{S},\text{BB}}$  resembles either the typical error or the spread of the flux densities within the BB. Furthermore, during this process, we preserved observational gaps ( $>60$  d periods with no data) by leaving them as empty periods in the BB light curve. We also ensured that all BBs are at least 1 d in length by iteratively merging and rebinning shorter BBs.

The BB algorithm operates using only one input parameter: a Bayesian prior called `nbp_prior`<sup>25</sup>. As `nbp_prior`  $\rightarrow \infty$ , the algorithm finds fewer BB; conversely, when `nbp_prior`  $\rightarrow 0$ , more BBs are found. For time series with ideal Gaussian errors, it is empirically shown that `nbp_prior`<sub>ideal</sub> =  $1.32 + 0.577 \log_{10}(N)$  where  $N$  is the total data count (Scargle et al. 2013). Unfortunately, this ideal assumption is not applicable to the merged CAZ light curves as they suffer from noise and non-intrinsic variability, especially due to the inclusion of ATLAS data (as shown in Sect. 4.1).

We minimized the effect of noise by calibrating `nbp_prior` using merged WD light curves. This process is described in more detail in Appendix B. When applying and calibrating the BB algorithm, we treated the CRTS light curves (i.e., the [C]V data) separately. This is because CRTS and other filters overlap minimally (if at all), resulting in two segments with vastly different cadence and, consequently, distinct `nbp_prior`<sub>ideal</sub>. We find that increasing the `nbp_prior`<sub>ideal</sub> of CRTS WD light curves by 0.4 and other WD light curves by 3.4 results in mostly constant BB light curves (see Appendix B). Thus, we adopt these enlargements to `nbp_prior`<sub>ideal</sub> when applying the BB algorithm to the CRTS and non-CRTS light curves of the CAZ catalog. Nevertheless, we caution that the BBs are still affected by extreme flux density outliers.

Due to the minimal overlap between the CRTS and non-CRTS segments, the CAZ light curves are often in two distinct segments that cannot be merged together. The first constitutes the CRTS data, whereas the second segment contains all other data (often merged together). In 5102 (66.1% of 7722) light curves, the CRTS and non-CRTS segments are separate, while they are

<sup>25</sup> Another parameter `gamma` is often used as the Bayesian prior, which follows the relation `nbp_prior` =  $-\ln(\text{gamma})$ .



**Fig. 7.** Distribution of the duration of BB95 in variable CAZ light curves. The vertical solid line shows the median of the distribution, and the dotted lines show its 25th and 75th percentiles.

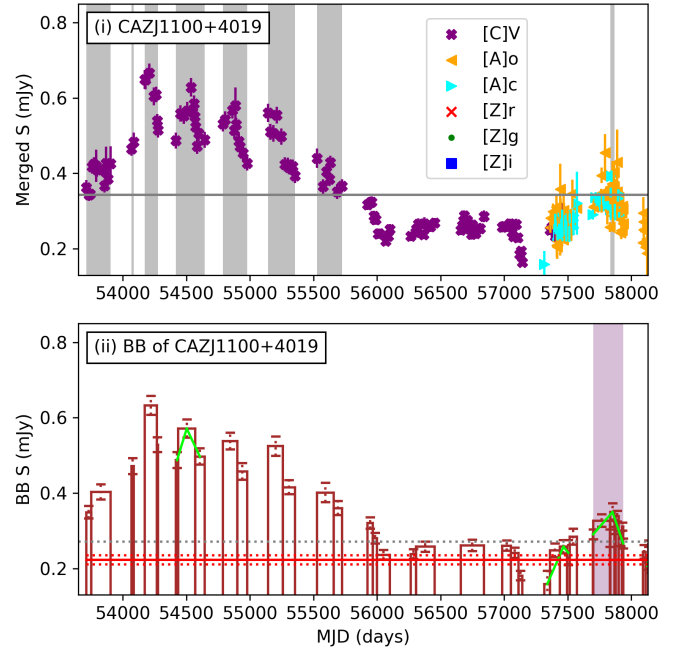
merged together in the remaining 2620 (33.9% of 7722). However, as mentioned above, the BBs calculated for the CRTS and non-CRTS segments need to be distinct regardless of their merge status. We combine the edges of the CRTS and non-CRTS BBs to obtain a single set of BBs in those 2620 light curves where they are merged together.

We identify periods of enhanced optical emission by finding BBs whose flux density is greater than the 95th percentile flux density (i.e.,  $S_{95\%}$ ), referred to as “BB95”. In the 5102 light curves where CRTS and non-CRTS segments are not merged together, we calculate their  $S_{95\%}$  separately. For the other 2620,  $S_{95\%}$  is calculated globally across the entire CAZ light curves. The BB95 periods are generally reliable and correspond to peaks of the most prominent flaring periods in a light curve. Nonetheless, when applied to CAZ light curves, they suffer from a few caveats: (1) they are easily affected by outliers (e.g., see panel i in Figs. 4 and 5 where a dim outlier with a flux density of  $\sim 0.7$  mJy at MJD 59 143 causes a BB to be not critically bright); (2) they can be unreliable in poorly shifted light curves; and (3) they exclude most periods of flaring activity since they only focus on BBs brighter than  $S_{95\%}$ . In Fig. 5 left panels, the gray vertical areas show BB95 in a few example light curves.

In this study we only focus on BB95 periods of confidently variable light curves (see Sect. 4.1). In total, we find 42754 BB95 across 2769 of 2798 variable sources. The reason for the lack of any BB95 in 29 variable sources is large flux density errors, which cause the binned flux densities to become dimmer than the global 95th percentile. In Fig. 7 we show the distribution of the BB95 durations, which are typically  $\sim 2$ –13 d. We note that a small peak at duration of  $10^{-2.69} = 0.002$  d arises by construction. If a single data point is preceded and proceeded by observational gaps, we define its BB duration as  $2 \times 0.001$  d. This is most common in the CRTS epoch.

#### 4.3. Identifying flaring periods

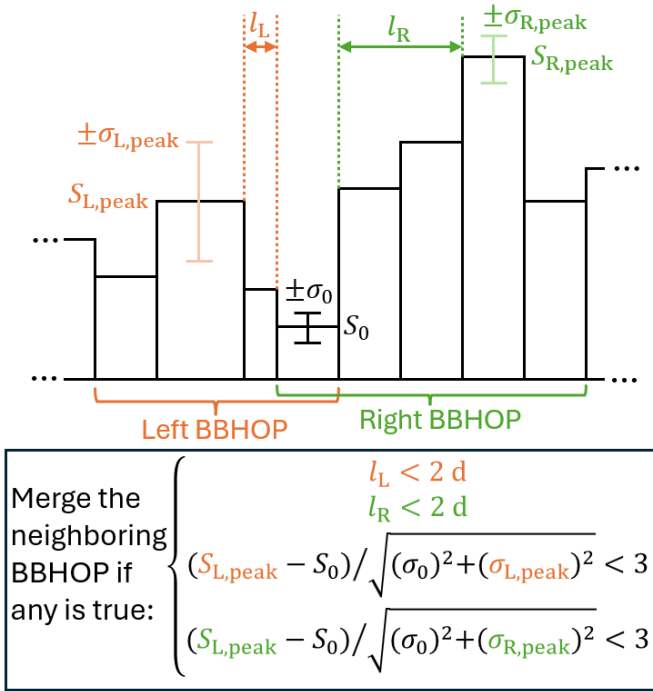
In this section we use BB differently by incorporating up- and down-going trends in flux density to more generally identify flaring periods in merged CAZ light curves. We apply a hopping algorithm (e.g., HOP, Eisenstein & Hut 1998) to find “hills” and “valleys” in the BB light curves. A hill is identified as a BB whose neighboring BBs (both rising and falling) are successively dimmer, and a valley as one whose neighboring BBs are successively brighter. Hills and valleys do not extend over observational gaps (see Sect. 4.2) of longer than 60 d. We select all groups of BBs from one valley to the next without going over the gaps. Using this hopping algorithm, each selected group starts on a valley, peaks on a hill, and terminates on a valley. We define each of these valley-hill-valley groups as a “BBHOP” flare (e.g., Meyer et al. 2019). The final BBHOP flares corresponding to the CAZ light curves are found in the electronic tables.



**Fig. 8.** Difference between BB95 (gray vertical areas in panel i), BBHOP flares (green lines in panel ii), and prominent BBHOP flares (purple area along with green lines in panel ii). Panel i: merged CAZ light curve. Panel ii: BBs of the merged CAZ light curve for a zoomed-in portion of CAZJ1100+4019. The horizontal solid gray line represents  $S_{95\%}$ , the horizontal dotted gray line represents  $S_{75\%}$ , and the horizontal solid red line shows  $S_{25\%}$  (with its errors shown as dotted red lines). For more plot details, see the description of Fig. 5. The BBHOP algorithm fails to find flares in low cadence periods (e.g., CRTS-epoch) due to frequent observational gaps. Notably, unlike the BBHOP algorithm, the performance of the BB95 algorithm on periods with data is not affected by the presence of neighboring observational gaps.

While the BBHOP algorithm identifies the majority of flaring periods, its reliability is limited by the cadence of the data. In epochs with low data count and frequent observational gaps (e.g., the CRTS-epoch of the merged CAZ light curves), the BBHOP algorithm cannot easily find valley-hill-valley BBs and thus BBHOP flares. This is not a limitation for the BB95 algorithm (see Sect. 4.2), thus offering an advantage over the BBHOP algorithm in such low cadence periods (e.g., Fig. 8), even though BBHOP flares do in general offer a more complete picture of the flaring activity.

Despite increasing the  $n_{cp\_prior\_ideal}$  (see Appendix B), the BB algorithm remains rather sensitive to fast variations in the merged CAZ light curves, especially in the more noisy ATLAS+ZTF-epoch. Fast variations can prematurely break valley-hill-valley BB patterns. This becomes detrimental when the BBHOP algorithm is tracing major outburst activity over longer rise and fall times. Moreover, blazars commonly show small flux variations during major outbursts, which also break the BBHOP continuity when tracing the underlying flare. To ensure that our BBHOP flares generally correspond to major flaring periods, we merge two BBHOP flares if: (1) the time difference from the peak BB to end BB of the first one is faster than 2 d; (2) the flux density decay of the first one is less significant than  $3\sigma$ ; (3) the time difference from the start BB to peak BB of the second one is faster than 2 d; or (4) the flux density rise of the second one is less significant than  $3\sigma$ . These criteria are visualized in Fig. 9. When two BBHOP flares are merged, they form a single BBHOP flare: the start BB of the first one becomes



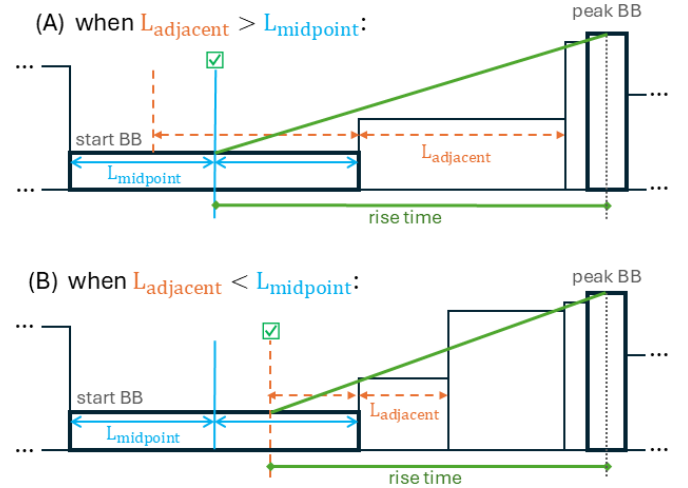
**Fig. 9.** Merging two neighboring BBHOP flares. If any of the four conditions is true, the two neighboring BBHOP flares are merged into a single BBHOP flare. The start BB, end BB, and peak BB of the merged flare are the start BB of the left BBHOP, end BB of the right BBHOP, and the larger of the two peak BBs, respectively.

the start BB of the merged flare, the end BB of the second one the end BB of the merged flare, and the larger of the two peak BBs is chosen as the peak BB of the merged flare. We iteratively merge all neighboring BBHOP flares (including newly merged ones with their potential neighbor) until neighboring BBHOP flares cannot be merged. Henceforth, the term “BBHOP flare” also refers to merged BBHOP flares.

The peak time of a BBHOP flare is defined as the middle of the peak BB. The start and end times of a flare are chosen using one of two methods. In the first method, the flare starts at the midpoint of the start BB. In the second method, we identify the BB adjacent to the start BB and use its duration to trace back in time from the start BB to obtain a start time for the flare (i.e., the adjacent block is “flipped”; e.g., Wagner et al. 2022). Between these two start times, we choose the one that minimizes the total duration of a flare. The end time of a flare is determined following the same logic, but mirrored. This minimization ensures that the flares do not have unnaturally long start or end times when they are preceded or succeeded by an unusually long quiescent period, respectively. These two scenarios are respectively sketched in panels A and B of Fig. 10.

We define “brightening” ( $\mathcal{B}$ ) as the significance of flux density change<sup>26</sup> of the start BB to the peak, and “decaying” ( $\mathcal{D}$ ) as that of the end BB to the peak. In addition to these significances of local flux density change, we define  $\mathcal{G}$  as the significance of flux density change from a global quiescent-level to the peak. The global quiescent-level is defined as the median of the lowest 50% of the merged flux densities (i.e.,  $S_{25\%}$ ). In this work, we refer to  $\mathcal{B}$ ,  $\mathcal{D}$ , and  $\mathcal{G}$  as flare “amplitudes”, which should not be confused with the convention where the flare amplitude is defined as its flux density change. Nonetheless,  $\mathcal{B}$ ,  $\mathcal{D}$ , and  $\mathcal{G}$

<sup>26</sup> Significance of change from  $S_1 \pm \sigma_1$  to  $S_2 \pm \sigma_2$  is  $\frac{S_2 - S_1}{\sqrt{\sigma_1^2 + \sigma_2^2}}$ .



**Fig. 10.** Scenarios used to determine the start time of a BBHOP flare. In scenario (A), the midpoint of the start BB is chosen as the start, as it minimizes the flare duration. Conversely, in scenario (B), the duration of the adjacent BB is used to obtain the start, as flipping it minimizes the flare duration. A mirrored equivalent of this principle is used for determining the fall time of a flare.

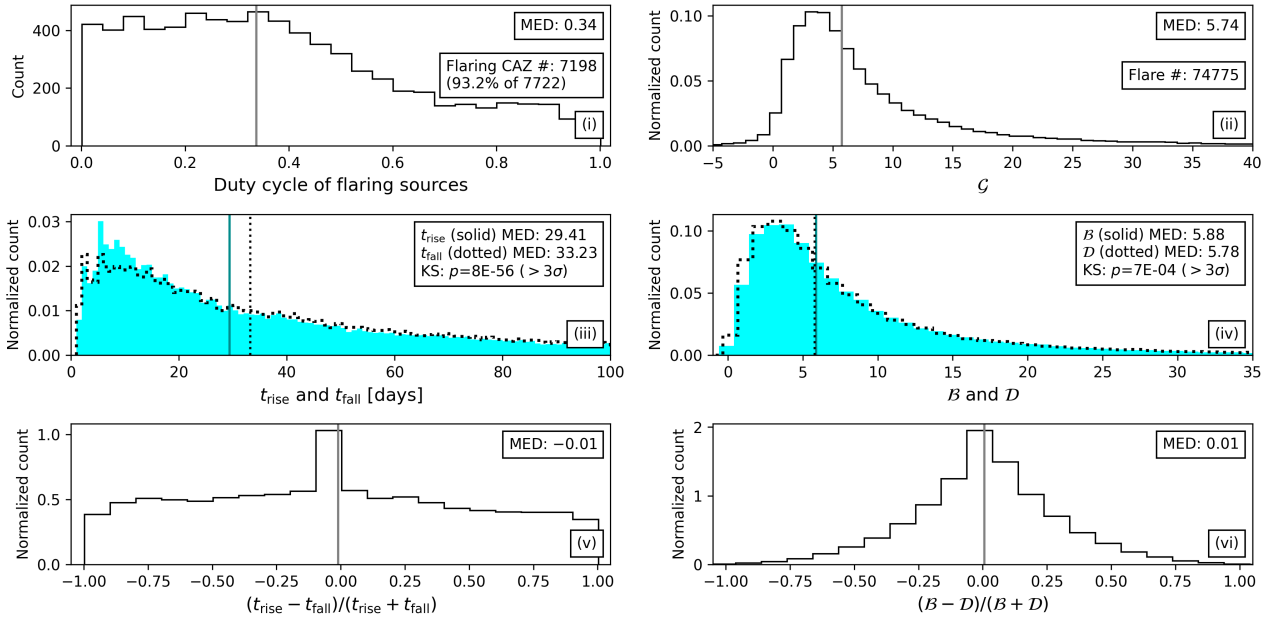
are directly proportional to flux density change, but additionally depend on the flux density errors.

We defined  $t_{\text{rise}}$  as the time taken between the start and peak of a flare (see Fig. 10), and  $t_{\text{fall}}$  between its peak and end. We investigate the asymmetry in the rise and fall times using  $(t_{\text{rise}} - t_{\text{fall}})/(t_{\text{rise}} + t_{\text{fall}})$ . Likewise, we inspect for asymmetry in the local brightening and decaying amplitudes using  $(\mathcal{B} - \mathcal{D})/(\mathcal{B} + \mathcal{D})$ . We also define the “duty cycle” of a source as the ratio of the total flaring time to the total duration of all BBs in a light curve. The duration of gaps in the BB light curve is omitted from the duty cycle calculation.

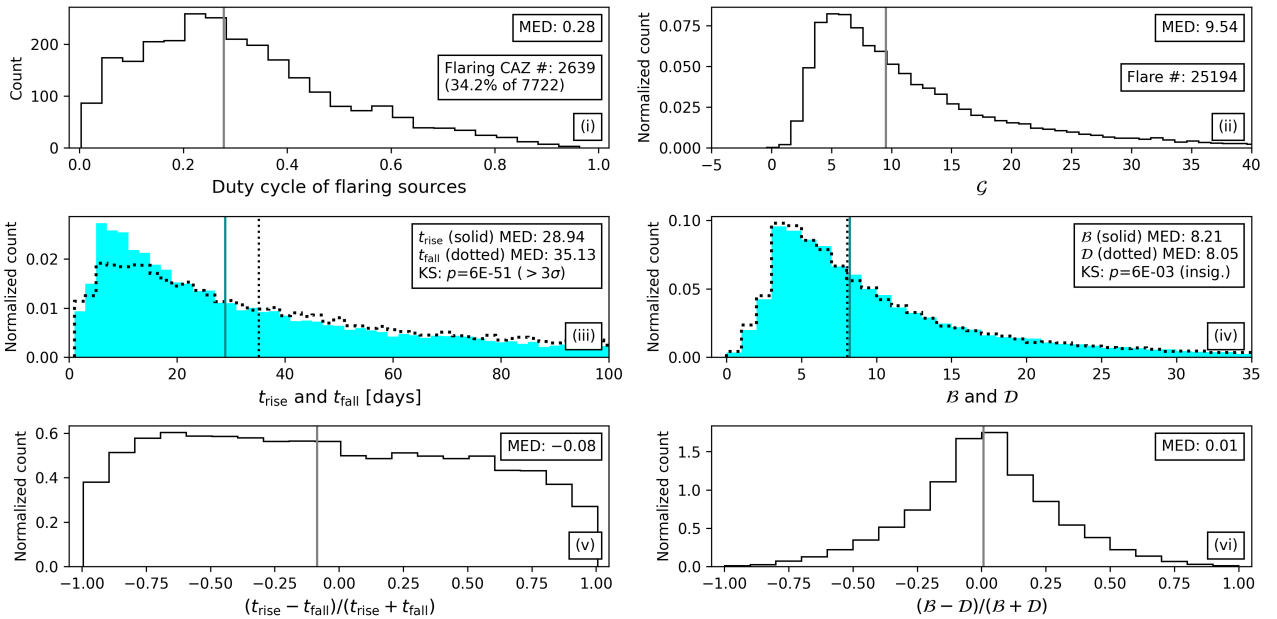
We looked for BBHOP flares in 7722 (97.5% of 7918 AGN) merged light curves with  $\geq 10$  data points and  $\geq 30$  d duration. We find 74775 BBHOP flares across 7198 (93.2% of 7722) light curves. In Fig. 11 we show the global distributions of the duty cycle,  $\mathcal{G}$ ,  $t_{\text{rise}}$  and  $t_{\text{fall}}$ ,  $\mathcal{B}$  and  $\mathcal{D}$ , as well as the temporal and amplitude asymmetry parameters.

Despite calibrating the Bayesian prior of the BB algorithm using WDS, we find that many BBHOP flares arise due to noise-driven scatter (e.g., see Fig. 5 where we show a zoomed-in version of Fig. 4). This is because: (1) the automatic forced-photometry and reduction algorithms of the all-sky surveys (especially those of ATLAS) cannot always reliably quantify the flux density and flux density error of blazars, particularly those which are host-galaxy-dominated<sup>27</sup>, resulting in extreme outliers; (2) the merged light curves combine data from three such surveys, with differing observational properties (e.g., limiting magnitudes, error estimations); and (3) blazars often exhibit

<sup>27</sup> Correcting for the effect of host-galaxy contamination on a blazar light curve requires meticulous modeling of the brightness of the host galaxy itself (e.g., Nilsson et al. 2007). This necessitates carefully centering an appropriately sized photometry aperture on the source, a process which is also sensitive to atmospheric conditions (i.e., the “seeing”). Additionally, the aperture size depends on the intrinsic brightness and angular size of the host galaxy, which should be determined on a source-by-source basis. Such careful host-galaxy modeling is not possible when using forced-photometry data and thus lies beyond the scope of this work.



**Fig. 11.** Flare characteristics of all 74775 BBHOP flares. The panels are as follows: (i) Distribution of the duty cycle of flaring sources. (ii) Global flare amplitudes ( $\mathcal{G}$ ). (iii) Rise and fall times of flares ( $t_{\text{rise}}$  and  $t_{\text{fall}}$ ). (iv) Local brightening and decaying flare amplitudes ( $\mathcal{B}$  and  $\mathcal{D}$ ). (v) Temporal asymmetry of flares. (vi) Amplitude asymmetry of flares. In panel iii we perform a KS test on the distributions of  $t_{\text{rise}}$  and  $t_{\text{fall}}$  and do the same on the distributions of  $\mathcal{B}$  and  $\mathcal{D}$  in panel iv. The term “MED” refers to the median of each distribution and is visualized using a vertical line.



**Fig. 12.** Flare characteristics of 25194 prominent BBHOP flares. For plot details, see the description of Fig. 11.

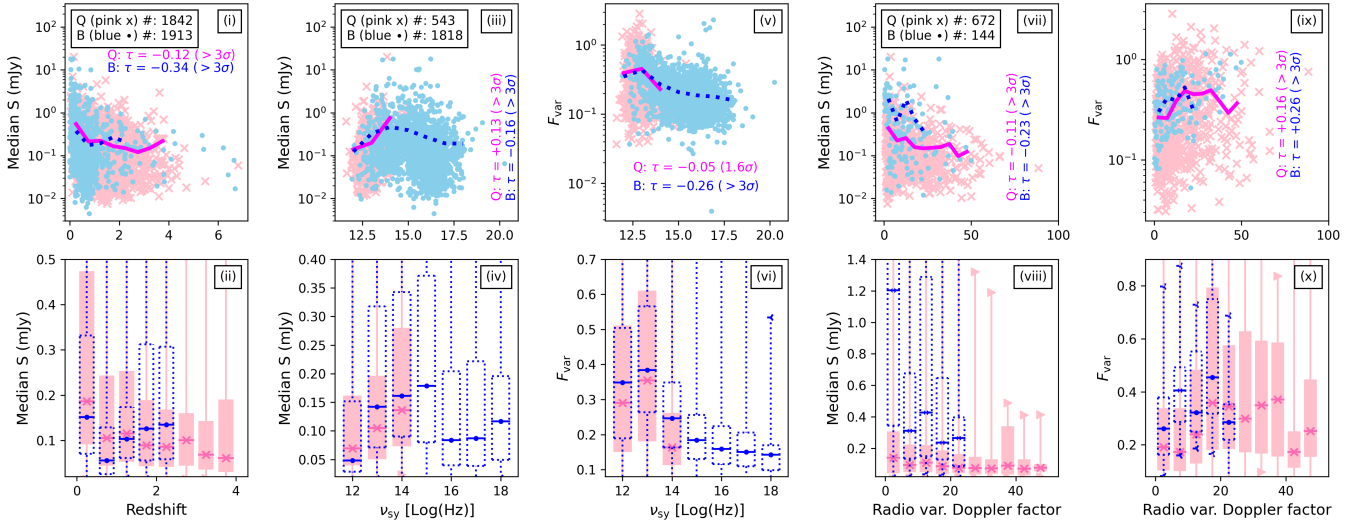
complex color variations and intrinsic properties across different filters which we could not account for when merging them. Therefore, when investigating the flaring properties of the two main blazar populations (FSRQs and BLLs) with respect to some physical properties (see Sect. 5.3), we mainly focus on the most prominent BBHOP flares.

We selected a prominent flare by requiring that (1) its light curve is variable (2798 of 7722 are; see Sect. 4.1), (2) its peak BB is brighter than the 75th percentile flux density (i.e.,  $S_{75\%}$ ), (3) one of its local amplitudes (i.e.,  $\mathcal{B}$  or  $\mathcal{D}$ ) is larger than 3, and (4) it is not a 3-block flare with only one data point in its peak BB. The first condition excludes 34 780 (46.5% of 74 775)

flares, the second another 11 333 (15.2%), the third another 2145 (2.9%), and the fourth another 1323 (1.8%). Thus, in total there are 25194 (33.7%) prominent BBHOP flares, whose global distributions are shown in Fig. 12. In Sect. 5.3 we discuss and interpret Figs. 11 and 12. Examples of BBHOP and prominent BBHOP flares are shown in Figs. 5 and 8.

## 5. Results and discussion

Here we describe and discuss the results of our variability and flare analyses of the CAZ light curves. In some parts, we only



**Fig. 13.** Distribution of median flux density against redshift (i and ii), median flux density and  $F_{\text{var}}$  against synchrotron peak frequency (iii–vi), and median flux density and  $F_{\text{var}}$  against radio variability Doppler factor (vii–x) in ZTF-only merged light curves of the sources in the CAZ catalog. Pink crosses and solid markings refer to FSRQs (Q); blue dots and dotted markings refer to BLLs (B). The top panels are scatter plots (with y-axis in logarithmic scales), where the binned running average is represented using lines (solid for FSRQs and dotted for BLLs). Kendall’s  $\tau$  correlation coefficient and significance are reported for FSRQs and BLLs. In the bottom panels, we show the binned box plots of each scatter plot. The averages and box plots are both only shown for bins with at least ten data points. In the top panels we show all data points, while in the bottom panels we zoom in on the boxes.

focus on the two main blazar subpopulations, FSRQs and BLLs (see Sect. 1), of which we have 2495 and 2726 sources, respectively (see Sect. 2). In Sect. 5.1 their overall variability via  $F_{\text{var}}$  is investigated and interpreted, in Sect. 5.2 their periods of enhanced emission via BB95, and in Sect. 5.3 their flaring behavior via BBHOP.

### 5.1. Variability of CAZ light curves

We measure the overall variability of a light curve using  $F_{\text{var}}$ . As mentioned in Sect. 4.1, the inclusion of the ATLAS data in the merged CAZ light curves introduces non-intrinsic variability. Thus, we investigate the behavior of blazar variability with respect to synchrotron peak frequency and radio variability Doppler factor only using survey-specific (CRTS- and ZTF-only) merged light curves. Since these distributions for CRTS- and ZTF-only light curves are comparable and lead to identical conclusions, for brevity, we just present the results for ZTF-only light curves.

In addition to  $F_{\text{var}}$ , we also investigate the behavior of median flux density against the above physical parameters as well as redshift. A caveat associated with this analysis is the shifted nature of the filters (see Sect. 4). Even when we only focus on ZTF-only merged light curves, some shifting issues are present since [Z]r, [Z]g, and [Z]i are shifted onto each other (in order of decreasing priority; see Sect. 3.3).

#### 5.1.1. Flux density against redshift

In Fig. 13i and ii, and all such plots in this section, we show the distribution of median flux density of the two main blazar populations (FSRQs and BLLs) against redshift ( $z$ ). In panel i, we calculate Kendall’s  $\tau$  non-parametric correlation coefficient for the FSRQ and BLL scatter plots separately.  $\tau$  ranges from +1 to 0 then to  $-1$ , indicating a perfect positive correlation, no correlation, and a perfect negative correlation, respectively. The significance of  $\tau$  is given in parentheses in terms of Gaussian  $\sigma$ . In panel ii, we draw box plots and zoom in. A box extends

from the 25th percentile to the 75th percentile of the binned data, with the median (50th percentile) being marked in between them. The lines below and above the box extend until the minimum and maximum values, respectively. In the following discussions we generally focus on the behavior of the boxes rather than the extensions. We note that the box plots as well as the running averages in the scatter plots are shown only when there are at least ten data points in each bin.

As seen from Fig. 13 (i and ii), we find an overall decreasing trend in blazar brightness as redshift increases as evident from the  $>3\sigma$  mildly negative  $\tau$  correlations of  $-0.12$  and  $-0.34$  for FSRQs and BLLs, respectively. Interestingly though, this decrease in brightness flattens after  $z \sim 0.5$ , similar to Abrahamyan et al. (2019). This flattening is also seen when only the more reliable spectroscopic redshifts are used. This apparent disagreement with the inverse-square law dimming is not unexpected as many factors affect blazar luminosity, namely: an observation bias toward greater relativistic boosting at larger redshifts (e.g., Lister & Marscher 1997; Hovatta et al. 2009; also see Sect. 5.1.3), and an increase in AGN density and activity at  $z \sim 2$  (e.g., Hopkins et al. 2007). The trend is even more complicated in the optical band. Firstly, at lower redshifts the effect of host-galaxy contribution is more prominent (especially in BLLs; e.g., Kotilainen et al. 1998). We tested this using the methodology of Tachibana & Miller (2018), confirming that a larger fraction of our sources at  $z \lesssim 0.5$  are galaxy-like. This is partly why we have brighter sources at  $z \lesssim 0.5$ . Secondly, the behavior of optical brightness and variability of a blazar against redshift depends strongly on its SED class and shape (i.e., spectral index), because at higher redshifts the observed photons originate from an increasingly higher frequency part of the SED (see Sect. 5.1.2). Accounting for this effect requires K-correction (e.g., Richards et al. 2006) which cannot be coherently performed for such a large number of multi-filter light curves and is thus beyond the scope of this work. Therefore, in this paper we do not investigate variability and flare characteristics of the CAZ light curves against redshift.

### 5.1.2. Flux density and variability against synchrotron peak frequency

Figure 13 (iii–vi) shows the median flux density and  $F_{\text{var}}$  of ZTF-only merged light curves against synchrotron peak frequency. As expected, higher synchrotron peak frequencies (ISP and HSP regimes) are only occupied by BLLs, since FSRQs are generally LSPs whereas BLLs can range from LSPs to EHSPs (Sect. 1).

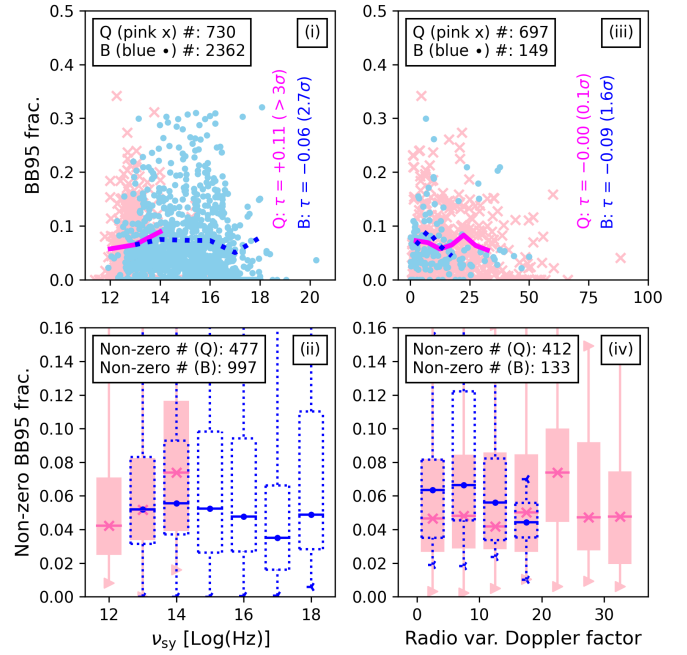
Emissions originating from closer to the synchrotron peak are brighter and more variable, due to the electrons having the highest energies, cooling the fastest, and being closest to the acceleration zones. This leads to a universal decrease in brightness and variability as lower and higher frequencies than  $\nu_{\text{sy}}$  are probed (e.g., Ikejiri et al. 2011; Ackermann et al. 2011; Hovatta et al. 2014; Richards et al. 2014; Baloković et al. 2016; Itoh et al. 2016; Safna et al. 2020; Otero-Santos et al. 2023). This is what we see in Fig. 13, where both FSRQs and BLLs peak in flux density at  $\nu_{\text{sy}} \sim 10^{14}$  Hz, and in  $F_{\text{var}}$  at  $\nu_{\text{sy}} \sim 10^{13}$  Hz, both of which are generally comparable to the optical band frequency of  $\sim 10^{14}$  Hz. We note that these trends are supported by the significant ( $>3\sigma$ )  $\tau$  correlations present for FSRQs and BLLs in their respective  $\nu_{\text{sy}}$  range. Interestingly, FSRQs and BLLs generally behave the same at the same  $\nu_{\text{sy}}$  range, in agreement with Hovatta et al. (2014).

### 5.1.3. Flux density and variability against radio variability Doppler factor

In Fig. 13 (vii–x) we explore how median flux density and  $F_{\text{var}}$  behave in terms of radio variability Doppler factor, which is calculated using flare amplitudes in the radio band (e.g., Lioudakis et al. 2018a; also see Sect. 2). Doppler factor generally quantifies the effect of relativistic boosting and is obtained using  $\delta = [\Gamma(1 - \beta \cos(\theta))]^{-1}$  where  $\Gamma = (1 - \beta^2)^{-1/2}$  is the Lorentz factor,  $\beta$  the bulk plasma speed in  $c$ , and  $\theta$  the viewing angle (e.g., Blandford et al. 2019). Therefore, as the bulk plasma speed increases and as the viewing angle decreases, Doppler factor increases. An increased Doppler factor results in greater relativistic boosting of the observed luminosities and variability amplitudes (e.g., Blandford & Königl 1979). This leads to the observational bias of having a greater fraction of high Doppler factor blazars at larger redshifts.

Since a larger radio variability Doppler factor implies an enhancement of relativistic boosting due to physical (i.e., increased bulk Lorentz factor) and geometrical (i.e., smaller viewing angle) properties of the jet, emissions in other bands should also experience similar enhancements. We see a hint of this with the  $>3\sigma$  mildly positive  $\tau$  correlations of +0.16 and +0.26 for FSRQs and BLLs, respectively, suggesting that  $F_{\text{var}}$  increases as radio variability Doppler factor increases. Nonetheless, the increasing trend appears to flatten at Doppler factor  $\gtrsim 20$ , which may be due to lower source counts. Further evidence of this trend is provided in Sect. 5.3.3 using BBHOP flares.

Surprisingly, flux density shows a general drop as the radio variability Doppler factor increases ( $>3\sigma$  mildly negative  $\tau$  correlations of  $-0.11$  and  $-0.23$  for FSRQs and BLLs, respectively). This is likely due to increased host-galaxy contamination for sources with smaller Doppler factors as they are most often found at low redshifts. This effect is especially noticeable for BLLs. We note that this could also be a reason why  $F_{\text{var}}$  is lower at smaller Doppler factors, since host-galaxy dilution can suppress variability. Another reason for the decreasing trend of median flux density with Doppler factor could be due to the



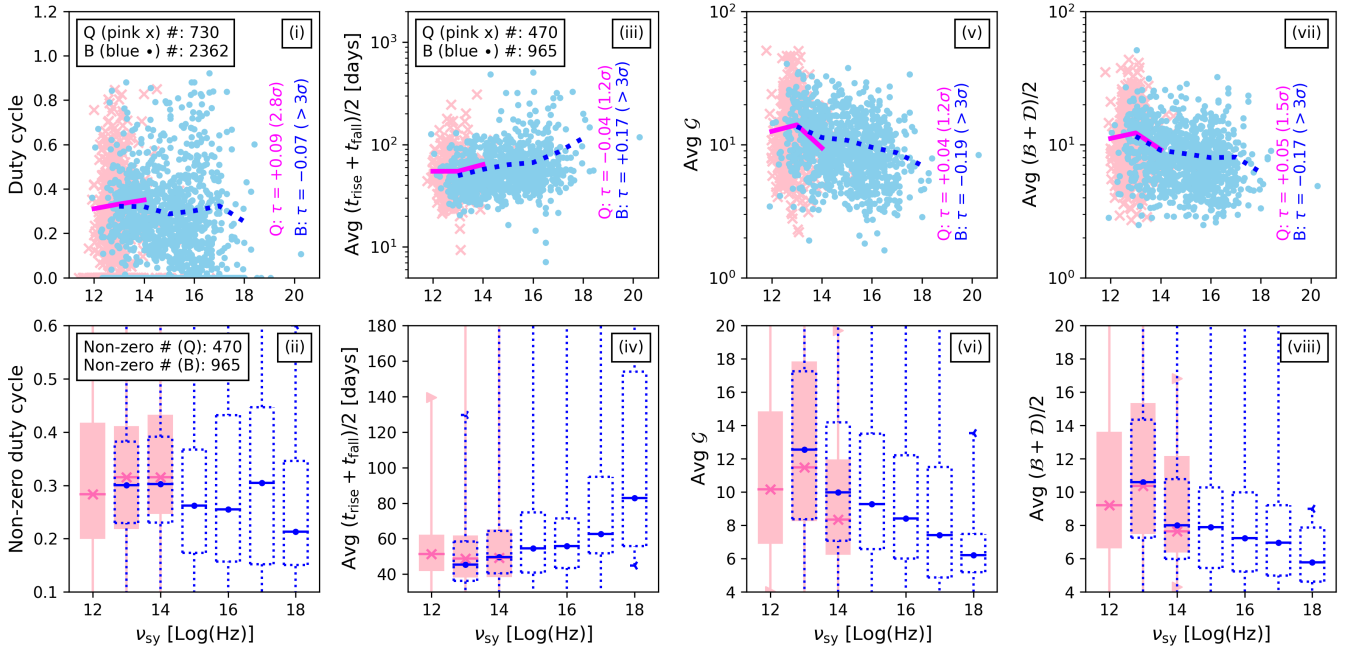
**Fig. 14.** Distribution of BB95 (with respect to all BB) fraction against synchrotron peak frequency (i and ii) and radio variability Doppler factor (iii and iv). Pink crosses and solid markings refer to FSRQs (Q); blue dots and dotted markings refer to BLLs (B). Kendall’s  $\tau$  correlation coefficient and significance are reported for FSRQs and BLLs. If a source has a BB95 fraction of zero, it is plotted in the top panels but not in the bottom ones. The running averages in the top row (shown as solid lines for FSRQs and dotted lines for BLLs) and the box plots in bottom row are both only calculated for at least ten nonzero values within each bin. In the top panels we show all data points, while in the bottom panels we zoom in on the boxes.

aforementioned observational bias. In other words, a larger fraction of high Doppler factor sources are expected to be at higher redshifts and thus dimmer in flux density.

### 5.2. Periods of enhanced emission in CAZ light curves

We plot the distribution of BB95 to all BB fraction against synchrotron peak frequency and radio variability Doppler factor in Fig. 14. We note that these BB95 are limited to those from variable CAZ light curves (see Sect. 4.1). The BB95 fraction of FSRQs shows a  $>3\sigma$  weakly positive trend ( $\tau = +0.11$ ) and peaks at  $\nu_{\text{sy}} \sim 10^{14}$ . For BLLs, the median of the BB95 fraction is also greatest at  $\nu_{\text{sy}} \sim 10^{14}$ . However, its trend is mostly flat against  $\nu_{\text{sy}}$  ( $\tau = -0.06$  at a significance of  $2.7\sigma$ ). As such, we cannot significantly determine whether the fraction of BB95 peaks at  $\nu_{\text{sy}} \sim 10^{14}$ . If it did, it would suggest that blazars whose synchrotron emission peaks at the optical energy exhibit more frequent periods of enhanced emission when probed at the optical band, which would fall inline with the reasoning of Sect. 5.1.2. Interestingly, a similar behavior is seen with BBHOP duty cycles against  $\nu_{\text{sy}}$  (see Fig. 15 panels i and ii). Nevertheless, we reiterate that this trend cannot be confirmed.

As seen in Fig. 14 (iii–iv), BB95 fraction does not seem to depend on radio variability Doppler factor. This is expected as an increased Doppler factor does not imply an increased frequency of flaring in blazars. We explore this further using the duty cycle of BBHOP flares in Sect. 5.3.1.1.



**Fig. 15.** Distribution of source duty cycle (i and ii) as well as source-averaged average rise and fall times (iii and iv), global amplitude (v and vi), and average local brightening and decaying amplitudes (vii and viii) of prominent BBHOP flares against synchrotron peak frequency. Pink crosses and solid markings refer to FSRQs (Q); blue dots and dotted markings refer to BLLs (B). Kendall's  $\tau$  correlation coefficient and significance are reported for FSRQs and BLLs. In panel i we show zero duty cycle data, whereas in the other panels we do not. The running averages in the top row (shown as solid lines for FSRQs and dotted lines for BLLs) and the box plots in bottom row are both only calculated for at least ten nonzero values within each bin. In the top panels we show all data points, while in the bottom panels we zoom in on the boxes.

### 5.3. Flaring behavior in CAZ light curves

We identify flares in merged CAZ light curves by employing the BBHOP algorithm (see Sect. 4.3). In total we find  $\sim 75\text{k}$  BBHOP flares across 7198 light curves. The global distributions of these flares are given in Fig. 11. As discussed in Sect. 4.3, many of these are noise-like and cannot always be confidently identified as intrinsic flares. As a result, we only focus on  $\sim 25\text{k}$  prominent flares across 2639 light curves to investigate the physical nature of blazar flares. The global distributions of the prominent BBHOP flares are shown in Fig. 12.

#### 5.3.1. Global BBHOP flare characteristics

Here we discuss the distributions of the duty cycle, amplitude, and rise and fall times of BBHOP flares, which are found in Figs. 11 and 12.

##### 5.3.1.1. Duty cycles of BBHOP flares

As evident from Fig. 11, the majority (93.2%) of CAZ light curves have at least one BBHOP flare. However, only 34.2% have at least one prominent flare (see Fig. 12). When limited to prominent flares, the duty cycle (fraction of flaring time to the total BB time) is generally small ( $\leq 0.3$ ), but can be as large as 0.9. Such a wide range of duty cycles is in agreement with Bauer et al. (2009), who studied optical variability of 362 blazars.

We note that large BBHOP duty cycles are possible because the entire duration of a BBHOP flare, including the low flux rising or falling portions, is used when calculating the total flaring time. Thus, light curves that exhibit back-to-back major flares with rather extended rise and fall times are likely to have very large BBHOP duty cycles. For the fraction of time CAZ light curves spend at high flux density, we refer the reader to BB95 fraction (see Sect. 5.2). Nevertheless, we caution that some large

duty cycles arise due to non-intrinsic flux density variations in problematic light curves. Despite our light curve cleaning efforts and stringent selection criteria for the prominent BBHOP flares, not all of them are intrinsic and their overall duration can become greatly exaggerated (especially as a result of BBHOP merging; see Sect. 4.3).

##### 5.3.1.2. Amplitudes of BBHOP flares

Global amplitude ( $\mathcal{G}$ , measured as the flux density significance change from the peak BB to a global quiescence level in Sect. 4.3) peaks at  $\sim 3$  for all flares (Fig. 11 plot ii) and  $\sim 5$  for prominent flares (Fig. 12 plot ii). Meanwhile, the local amplitudes (brightening  $\mathcal{B}$  and decaying  $\mathcal{D}$ , measured as the flux density significance change from the peak BB to the starting and ending BBs, respectively) peak at  $\sim 3.1$  for all flares (Fig. 11 plot iv) and at  $\sim 3.6$  for prominent ones (Fig. 12 plot iv). These trends are expected based on the selection criteria of the prominent flares.

Regarding amplitude asymmetry, as seen in plot (vi) of Figs. 11 and 12 via the  $(\mathcal{B} - \mathcal{D})/(\mathcal{B} + \mathcal{D})$  distributions and the two-sample Kolmogorov-Smirnov (KS) tests,  $\mathcal{B}$  and  $\mathcal{D}$  are generally symmetrical (the  $>3\sigma$  KS test becomes insignificant when going from all to prominent flares). This intuitive result indicates that flaring states in blazars are transient and all blazars eventually return to a quiescent state.

##### 5.3.1.3. Timescales of BBHOP flares

The rise and fall times ( $t_{\text{rise}}$  and  $t_{\text{fall}}$ , respectively) of BBHOP flares peak at  $\sim 7$  d. These timescales are generally comparable the  $\sim 1\text{--}8$  d characteristic timescales found by Dingler & Smith (2024) for 67 blazars. The rise and fall times exhibit a substantial spread, with their 25th and 75th percentiles being  $\sim 13\text{--}63$  d and  $\sim 16\text{--}71$  d, respectively. This translates to typical BBHOP flare durations of  $\sim 30\text{--}130$  d. The large spread is compatible with the

characteristic timescales found by Xiong et al. (2025) for 1684 blazar ZTF light curves. Interestingly, BBHOP flares are clearly asymmetric in time (KS  $p$ -values of  $<10^{-50}$ , i.e., KS significance of  $\gg 3\sigma$ ) with a faster rise than fall. For example, the median rise time is  $\sim 29$  d while the median fall time  $\sim 35$  d. The distribution of  $(t_{\text{rise}} - t_{\text{fall}})/(t_{\text{rise}} + t_{\text{fall}})$  is skewed toward the negative values with a median of  $-0.08$ .

Blazars are known to show both temporally symmetrical (e.g., Chatterjee et al. 2012; Blinov et al. 2015; Cerruti 2025) and asymmetrical flares (e.g., Nalewajko 2013; Roy et al. 2019; Valtaoja et al. 1999; Nieppola et al. 2009) in multiple bands. Temporal symmetry generally arises when light crossing time dominates the shape of the flares, while asymmetry from rapid acceleration or cooling of the emitting particles. Both symmetrical and asymmetrical flares can be explained by blazar acceleration models (with shock recollimation models, e.g., Fromm et al. 2016; Fichet de Clairfontaine et al. 2021; magnetic reconnection, e.g., Cerutti et al. 2012; Zhang et al. 2017; and either of the two, e.g., Potter 2018). We do not closely compare the predictions of these models to our BBHOP flares, because such model comparisons require careful flare selection (e.g., Jormanainen 2018), which is beyond the scope of this work involving tens of thousands of flares.

### 5.3.2. BBHOP flare characteristics against synchrotron peak frequency

In Fig. 15, we show the duty cycle, source-averaged average rise and fall times (i.e., half of flare duration on average), source-averaged global amplitude, and source-averaged average local amplitudes of the prominent BBHOP flares against synchrotron peak frequency. In Appendix C we show these plots using all BBHOP flares (see Fig. C.1). The main trends and results are identical between Figs. 15 and C.1.

Regarding BBHOP duty cycles, as seen in Fig. 15 (i and ii), while a peak at  $\nu_{\text{sy}} \sim 10^{13}$  Hz may be present for duty cycle against  $\nu_{\text{sy}}$ , this cannot be confirmed. This is because the duty cycle is mostly flat, albeit very weakly dropping, as  $\nu_{\text{sy}}$  increases (for BLLs  $\tau = -0.07$  at  $>3\sigma$ ). This almost flat behavior is similar to that of the BB95 fraction against  $\nu_{\text{sy}}$  in Fig. 14 (see Sect. 5.2). Interestingly, the duty cycle of all BBHOP flares (Fig. C.1 panels i and ii) shows a clearer drop with increasing  $\nu_{\text{sy}}$  (for BLLs  $\tau = -0.18$  at  $>3\sigma$ ). Such a decreasing trend would generally agree with Paliya et al. (2017), who monitored the intranight optical variability of 17 blazars and found that LSPs are more variable and show higher duty cycles than ISPs and HSPs. Nonetheless, we cannot confidently claim a clear decreasing trend in duty cycle against  $\nu_{\text{sy}}$ .

From Fig. 15 (iii and iv), we see that the average rise and fall times for prominent flares is smallest at  $\nu_{\text{sy}} \sim 10^{13}$  Hz with a clear increasing trend toward higher  $\nu_{\text{sy}}$  (for BLLs  $\tau = +0.17$  at  $>3\sigma$ ). This agrees with the argument presented in Sect. 5.1.2, since probing lower energies than the synchrotron peak frequency means observing particles which have longer cooling times. Additionally, the flaring timescales for FSRQs and BLLs are on average comparable (regardless of  $\nu_{\text{sy}}$ ), which is similar to their behavior in the radio band (e.g., Hovatta et al. 2007).

Likewise, from Fig. 15 (v–viii), we see that the global and local amplitudes of the prominent BBHOP flares peak at  $\nu_{\text{sy}} \sim 10^{13}$  Hz and exhibit a clear drop toward higher  $\nu_{\text{sy}}$ , as evident by the  $>3\sigma$  mildly negative  $\tau$  correlations of  $-0.19$  and  $-0.17$  for the BLL global and local amplitudes, respectively, against  $\nu_{\text{sy}}$ . A similar behavior with respect to fractional variability was

seen in Sect. 5.1.2, where the highest variability occurs at the synchrotron peak frequency. This is also inline with the reasoning presented in Sect. 5.1.2, since emissions corresponding to the synchrotron peak frequency come from particles with the highest energies and thus exhibit the largest flares. Notably, even in terms of flare amplitudes, FSRQs and BLLs at the same  $\nu_{\text{sy}}$  range behave similarly.

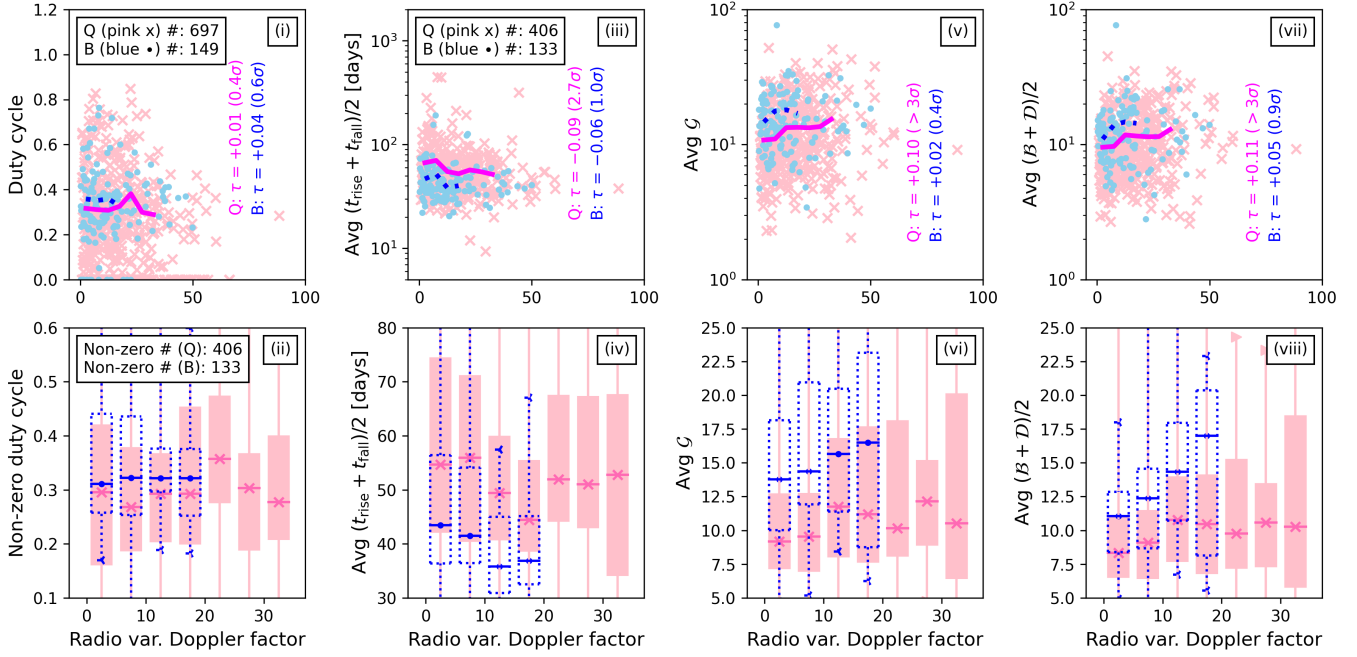
The similarities between the flux density, variability, and flaring behaviors of FSRQs and BLLs at the same  $\nu_{\text{sy}}$  range suggest that the processes which dominate their flaring behavior are likely similar. Crucially, this suggests that differences between the variability and flaring behaviors of FSRQs and BLLs in studies where their entire populations are considered (i.e., without limiting  $\nu_{\text{sy}}$ ; e.g., Bauer et al. 2009; Zhang et al. 2024) generally arise due to the distinct  $\nu_{\text{sy}}$  ranges of FSRQs and BLLs.

### 5.3.3. BBHOP flare characteristics against radio variability Doppler factor

In Fig. 16, we show the duty cycle, source-averaged average rise and fall times, source-averaged global amplitude, and source-averaged average local amplitudes of prominent BBHOP flares against radio variability Doppler factor. These plots for all BBHOP flares are given in Appendix C Fig. C.2. The main trends and results are identical between Figs. 16 and C.2.

In Fig. 14 (i and ii), we see that the duty cycle of the prominent flares remains constant with respect to radio variability Doppler factor (similar to the behavior of BB95 fraction against radio variability Doppler factor; see Fig. 14). This behavior is expected because the frequency of events that lead to flares in blazar jets is not expected to be modified by Doppler factor. Nevertheless, a greater average Doppler factor should imply shortened rise and fall times because of shortened photon arrival timescales in the observer frame (i.e.,  $\Delta t_{\text{obs}} \propto 1/\delta$ , where  $\Delta t_{\text{obs}}$  and  $\delta$  are the observed flaring times and the Doppler factor, respectively; see Eq. (D.1)). Likewise, a greater average Doppler factor should result in larger flare amplitudes because of enhanced flux density in the observer frame (i.e.,  $\Delta S_{\text{obs}} \propto \delta^{3-\alpha}$ , where  $S_{\text{obs}}$ ,  $\delta$ , and  $\alpha$  are the observed flux densities at a given frequency, Doppler factor, and spectral index as defined by  $S \propto \nu^\alpha$ , respectively; see Eq. (D.2)).

We see a hint of these trends in Fig. 16 (iii–viii). For FSRQs, which have three times more radio variability Doppler factor estimates than BLLs, we find that the average flaring times show a very weak negative correlation with radio variability Doppler factor ( $\tau = -0.09$  at  $2.7\sigma$ ), and the global and local amplitudes weak positive correlations ( $\tau \approx +0.10$  at  $>3\sigma$ ). In Appendix D we correct the flare parameters for the effect of redshift and, by extension, luminosity distance. After this correction, the expected flare parameter versus Doppler factor trends become stronger and significant ( $>3\sigma$ ), as seen in Fig. D.1. The  $>3\sigma$  negative correlation between redshift-corrected flare times and radio variability Doppler factor is milder for FSRQs with  $\tau = -0.18$  than BLLs with  $\tau = -0.31$ . Likewise, the  $>3\sigma$  positive correlation between redshift-corrected flare amplitudes and radio variability Doppler factor is milder for FSRQs with  $\tau = +0.19$  than BLLs with  $\tau \approx +0.40$ . Nevertheless, we caution that these trends may be partially due to host-galaxy contamination for low redshift sources which affects BLLs more strongly (see Appendix D for more details). It is noteworthy that such flare time and amplitude trends against variability Doppler factor were also seen by Hovatta et al. (2007) who analyzed radio flares in 53 blazars.



**Fig. 16.** Similar to Fig. 15 but for the radio variability Doppler factor. We show similar plots using redshift-corrected parameters in Appendix D.

## 6. Summary and conclusion

In this paper we have constructed a catalog of 7918 blazars and blazar candidates, which is the largest of its kind to date. This catalog is referred to as CAZ and was constructed by combining several blazar-dominated AGN catalogs and samples (RFC, 4LAC, 3HSP, CGRaBS, and 5BZC; Sect. 2). The electronic tables accompanying this paper contain the RA and Dec. of the sources and, when available, their source type, redshift, SED class, synchrotron and HE SED peak frequencies, radio variability Doppler factor, and median X-ray flux density (Sect. 2). Their X-ray light curves and spectra as well as an X-ray-selected, statistically complete subsample of the CAZ catalog are found in our companion paper, Paggi et al. (in prep.).

Additionally, in this paper we extracted optical light curves for sources in the CAZ catalog via three all-sky surveys: CRTS, ATLAS, and ZTF (Sect. 3). We cleaned and merged all the filters (Sects. 3.2 and 3.3), obtaining light curves that are as long and as dense as possible for as many of the 7918 sources as possible. The merged light curves as well as their corresponding BB data and BBHOP flares are given in the electronic tables.

Furthermore, we first quantified the overall variability of the CAZ light curves via  $F_{\text{var}}$  (see Sect. 4.1 for the analysis details, and Sect. 5.1 for the results). Then, we identified their periods of enhanced emission using critically bright BBs (see Sect. 4.2 for the analysis details, and Sect. 5.2 for the results). Subsequently, we identified their flaring periods using BBHOP (see Sect. 4.3 for the analysis details, and Sect. 5.3 for the results). For some analyses, we only focused on the two main subpopulations of blazars: FSRQs (2495 sources) and BLLs (2726 sources).

Despite the notable increase in the number of blazars and their available parameters, all of our results are generally in agreement with prior studies (as discussed in Sect. 5). Below is a summary of the results:

- Confidently variable CAZ light curves spend  $\sim 30\%$  of the time undergoing prominent flares ( $S_{\text{peak}} > S_{75\%}$ ), with each flare lasting  $\sim 30\text{--}130$  d.
- Confidently variable CAZ light curves spend  $\sim 5\%$  of the time with  $S > S_{95\%}$  for typical durations of  $\sim 2\text{--}13$  d.

- Optical flares generally have a faster rise than decay.
- Blazars are brightest and most variable – exhibiting higher amplitude and faster flares – when observed at frequencies close to their  $\nu_{\text{sy}}$ . Their brightness and variability decrease as lower and higher frequencies than  $\nu_{\text{sy}}$  are probed.
- Optical variability and flare characteristics in FSRQs and BLLs are comparable at the same  $\nu_{\text{sy}}$  range, hinting that their flaring processes are similar.
- Optical flare times tend to decrease while flare amplitudes increase as the radio variability Doppler factor increases.

The CAZ catalog and light curves allow blazar and, by extension, jetted AGN variability studies to be done with unprecedentedly large sample sizes, which can be vital for understanding various physical phenomena involving SMBH systems and their jets. For example, such long-term, high-cadence, and highly numerous light curves can be used to (1) gain insights into the physics of accretion and properties of the central engine (e.g., Kelly et al. 2009; Caplar et al. 2017; Sun et al. 2020); (2) investigate jet processes and infer the physical properties of the emitting regions (e.g., Carini et al. 2011; Jorstad et al. 2013; Marscher 2014; Jormanainen et al. 2023, 2025; Kankkunen et al. 2025b); (3) identify binary SMBH systems or dynamically moving jets (e.g., Sillanpaa et al. 1988; Raiteri et al. 2017; Lioudakis et al. 2021); (4) boost the statistical significance of population-based correlation studies (e.g., the elusive blazar-neutrino connection, Lioudakis et al. 2022a; Kouch et al. 2024a, 2025a, 2026); (5) understand the origin of the HE emission in blazars through multiwavelength cross-correlation analyses (e.g., Fuhrmann et al. 2014; Ramakrishnan et al. 2015; Lindfors et al. 2016; Lioudakis et al. 2018b; de Jaeger et al. 2023); (6) identify counterparts in higher energy observations (e.g., Abdo et al. 2010b; Ruan et al. 2012); (7) classify blazars (e.g., Żywucka et al. 2018; Agarwal 2023); (8) understand the role of AGN in shaping the evolution of their host galaxy (e.g., Hickox et al. 2014); and (9) prepare for next-generation transient data (e.g., Vera C. Rubin Observatory’s Legacy Survey of Space and Time; Ivezić et al. 2019).

## Data availability

The CAZ catalog containing 7918 sources along with their compiled physical parameters as well as the corresponding CAZ light curves, BB information, and BBHOP flare information are available at the CDS via <https://cdsarc.cds.unistra.fr/viz-bin/cat/J/A+A/708/A382>. The catalog provides the J2000 RA and Dec of the CAZ sources along with their RFC, 4LAC, 3HSP, CGRaBS, and 5BZC names as well as their source type, redshift, synchrotron and high-energy peak frequencies, radio variability Doppler factor, median X-ray flux density, fractional variability in the CRTS and ZTF portions of the CAZ light curve, their CAZ variability status, BB95 fraction, and prominent BBHOP duty cycle. The CAZ light curve of each source, if available, is given as MJD, merged flux density, flux density error, filter, shift factor, and merge information (i.e., which filters are merged together). The BB information of each source, if available, is given as MJD of the left and right edges of a BB, average flux density within the BB and its error, number of data points within the BB, and if the BB is considered a BB95. The BBHOP flare information of each source, if available, is given as the MJD of the start, peak, and end of a BBHOP flare along with the respective average flux density and flux density error at each of those MJDs, along with the information if the BBHOP flare is a prominent one.

*Acknowledgements.* We thank the anonymous referee for constructive comments. We also thank Matteo Cerruti for fruitful discussion on flare symmetry. P.K. was supported by the Research Council of Finland projects 346071 and 345899. E.L. was supported by the Research Council of Finland projects 317636, 320045, and 346071. TH was supported by the Research Council of Finland projects 317383, 320085, 345899, and 362571 and the European Union ERC-2024-COG - ARTICLES - 101169986. J.J. was supported by the Research Council of Finland projects 320085 and 345899. K.K. acknowledges support from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 101002352). I.L. and A.P. were funded by the European Union ERC-2022-STG - BOOTES - 101076343. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. Based on observations obtained with the Samuel Oschin Telescope 48-inch and the 60-inch Telescope at the Palomar Observatory as part of the *Zwicky* Transient Facility project. Z.T.F. is supported by the National Science Foundation under Grant No. AST-2034437 and a collaboration including Caltech, IPAC, the Weizmann Institute for Science, the Oskar Klein Center at Stockholm University, the University of Maryland, Deutsches Elektronen-Synchrotron and Humboldt University, the TANGO Consortium of Taiwan, the University of Wisconsin at Milwaukee, Trinity College Dublin, Lawrence Livermore National Laboratories, and IN2P3, France. Operations are conducted by COO, IPAC, and UW. The ZTF forced-photometry service was funded under the Heising-Simons Foundation grant #12540303 (PI: M.J.Graham). This work has made use of data from the Asteroid Terrestrial-impact Last Alert System (ATLAS) project. The Asteroid Terrestrial-impact Last Alert System (ATLAS) project is primarily funded to search for near earth asteroids through NASA grants NN12AR55G, 80NSSC18K0284, and 80NSSC18K1575; byproducts of the NEO search include images and catalogs from the survey area. This work was partially funded by Kepler/K2 grant J1944/80NSSC19K0112 and HST GO-15889, and STFC grants ST/T000198/1 and ST/S006109/1. The ATLAS science products have been made possible through the contributions of the University of Hawaii Institute for Astronomy, the Queen's University Belfast, the Space Telescope Science Institute, the South African Astronomical Observatory, and The Millennium Institute of Astrophysics (MAS), Chile. This work has made use of data from the Joan Oró Telescope (TJO) of the Montsec Observatory (OdM), which is owned by the Catalan Government and operated by the Institute for Space Studies of Catalonia (IEEC). This work makes use of Matplotlib (Hunter 2007), NumPy (Harris et al. 2020), SciPy (Virtanen et al. 2020), and Astropy (Astropy Collaboration 2022).

## References

Abdo, A. A., Ackermann, M., Agudo, I., et al. 2010a, *ApJ*, 716, 30

- Abdo, A. A., Ackermann, M., Ajello, M., et al. 2010b, *ApJ*, 722, 520  
 Abrahamyan, H. V., Mickaelian, A. M., Paronyan, G. M., & Mikavelyan, G. A. 2019, *Astron. Nachr.*, 340, 437  
 Ackermann, M., Ajello, M., Allafort, A., et al. 2011, *ApJ*, 743, 171  
 Agarwal, A. 2023, *ApJ*, 946, 109  
 Agudo, I., Lioudakis, I., Otero-Santos, J., et al. 2025, *ApJ*, 985, L15  
 Aharonian, F., Akhperjanian, A. G., Bazer-Bachi, A. R., et al. 2007, *ApJ*, 664, L71  
 Ajello, M., Baldini, L., Ballet, J., et al. 2022, *ApJS*, 263, 24  
 Albert, J., Aliu, E., Anderhub, H., et al. 2007, *ApJ*, 669, 862  
 Astropy Collaboration (Price-Whelan, A. M., et al.) 2022, *ApJ*, 935, 167  
 Atwood, W. B., Abdo, A. A., Ackermann, M., et al. 2009, *ApJ*, 697, 1071  
 Baloković, M., Paneque, D., Madejski, G., et al. 2016, *ApJ*, 819, 156  
 Bauer, A., Baltay, C., Coppi, P., et al. 2009, *ApJ*, 699, 1732  
 Bell, K. J. 2026, in *Encyclopedia of Astrophysics* (Amsterdam: Elsevier), 3, 75  
 Bellm, E. C., Kulkarni, S. R., Graham, M. J., et al. 2019, *PASP*, 131, 018002  
 Blandford, R., Meier, D., & Readhead, A. 2019, *ARA&A*, 57, 467  
 Blandford, R. D., & Königl, A. 1979, *ApJ*, 232, 34  
 Blinov, D., Pavlidou, V., Papadakis, I., et al. 2015, *MNRAS*, 453, 1669  
 Bonning, E., Urry, C. M., Bailyn, C., et al. 2012, *ApJ*, 756, 13  
 Böttcher, M. 2019, *Galaxies*, 7, 20  
 Caplar, N., Lilly, S. J., & Trakhtenbrot, B. 2017, *ApJ*, 834, 111  
 Carini, M. T., Walters, R., & Hopper, L. 2011, *AJ*, 141, 49  
 Cerruti, M. 2025, *A&A*, 698, A101  
 Cerruti, B., Werner, G. R., Uzdensky, D. A., & Begelman, M. C. 2012, *ApJ*, 754, L33  
 Chang, Y. L., Arsioli, B., Giommi, P., Padovani, P., & Brandt, C. H. 2019, *A&A*, 632, A77  
 Chatterjee, R., Bailyn, C. D., Bonning, E. W., et al. 2012, *ApJ*, 749, 191  
 Cohen, D. P., Romani, R. W., Filippenko, A. V., et al. 2014, *ApJ*, 797, 137  
 Costamante, L., Ghisellini, G., Giommi, P., et al. 2001, *A&A*, 371, 512  
 de Jaeger, T., Shappee, B. J., Kochanek, C. S., et al. 2023, *MNRAS*, 519, 6349  
 Dingler, R., & Smith, K. L. 2024, *ApJ*, 973, 10  
 Drake, A. J., Djorgovski, S. G., Mahabal, A., et al. 2009, *ApJ*, 696, 870  
 Eisenstein, D. J., & Hut, P. 1998, *ApJ*, 498, 137  
 Fichet de Clairfontaine, G., Meliani, Z., Zech, A., & Hervet, O. 2021, *A&A*, 647, A77  
 Filippenko, A. V., Li, W. D., Treffers, R. R., & Modjaz, M. 2001, *ASP Conf. Ser.*, 246, 121  
 Fiorucci, M., Ciprini, S., & Tosti, G. 2004, *A&A*, 419, 25  
 Fromm, C. M., Perucho, M., Mimica, P., & Ros, E. 2016, *A&A*, 588, A101  
 Fuhrmann, L., Larsson, S., Chiang, J., et al. 2014, *MNRAS*, 441, 1899  
 Gehrels, N., Chincarini, G., Giommi, P., et al. 2004, *ApJ*, 611, 1005  
 Ghisellini, G., Tavecchio, F., Foschini, L., et al. 2010, *MNRAS*, 402, 497  
 Ghisellini, G., Righi, C., Costamante, L., & Tavecchio, F. 2017, *MNRAS*, 469, 255  
 Glorioso, M., Sambruna, R. M., & Eracleous, M. 2003, *ApJ*, 584, 176  
 Gordon, D., Jacobs, C., Beasley, A., et al. 2016, *AJ*, 151, 154  
 Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Nature*, 585, 357  
 Healey, S. E., Romani, R. W., Cotter, G., et al. 2008, *ApJS*, 175, 97  
 Hickox, R. C., Mullaney, J. R., Alexander, D. M., et al. 2014, *ApJ*, 782, 9  
 Hopkins, P. F., Richards, G. T., & Hernquist, L. 2007, *ApJ*, 654, 731  
 Hovatta, T., & Lindfors, E. 2019, *New Astron. Rev.*, 87, 101541  
 Hovatta, T., Tornikoski, M., Lainela, M., et al. 2007, *A&A*, 469, 899  
 Hovatta, T., Valtaoja, E., Tornikoski, M., & Lähteenmäki, A. 2009, *A&A*, 494, 527  
 Hovatta, T., Pavlidou, V., King, O. G., et al. 2014, *MNRAS*, 439, 690  
 Hovatta, T., Lindfors, E., Kiehlmann, S., et al. 2021, *A&A*, 650, A83  
 Hunter, J. D. 2007, *Comput. Sci. Eng.*, 9, 90  
 Ikejiri, Y., Uemura, M., Sasada, M., et al. 2011, *PASJ*, 63, 639  
 Itoh, R., Nalewajko, K., Fukazawa, Y., et al. 2016, *ApJ*, 833, 77  
 Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, *ApJ*, 873, 111  
 Jansen, F., Lumb, D., Altieri, B., et al. 2001, *A&A*, 365, L1  
 Jormanainen, J. 2018, Master's thesis, University of Turku, Finland  
 Jormanainen, J., Hovatta, T., Christie, I. M., et al. 2023, *A&A*, 678, A140  
 Jormanainen, J., Hovatta, T., Lindfors, E., et al. 2025, *A&A*, 694, A206  
 Jorstad, S. G., Marscher, A. P., Smith, P. S., et al. 2013, *ApJ*, 773, 147  
 Jorstad, S. G., Marscher, A. P., Raiteri, C. M., et al. 2022, *Nature*, 609, 265  
 Kankkunen, S., Tornikoski, M., & Hovatta, T. 2025a, *A&A*, 693, A319  
 Kankkunen, S., Tornikoski, M., Hovatta, T., & Lähteenmäki, A. 2025b, *A&A*, 693, A318  
 Kelly, B. C., Bechtold, J., & Siemiginowska, A. 2009, *ApJ*, 698, 895  
 Kembhavi, A. K., & Narlikar, J. V. 1999, *Quasars and Active Galactic Nuclei: an Introduction* (Cambridge: Cambridge University Press)  
 Kotilainen, J. K., Falomo, R., & Scarpa, R. 1998, *A&A*, 336, 479  
 Kouch, P. M., Lindfors, E., Hovatta, T., et al. 2024a, *A&A*, 690, A111  
 Kouch, P. M., Lioudakis, I., Middei, R., et al. 2024b, *A&A*, 689, A119

- Kouch, P. M., Hovatta, T., Lindfors, E., et al. 2026, *A&A*, 708, A383
- Kouch, P. M., Lindfors, E., Hovatta, T., et al. 2025a, *A&A*, 696, A73
- Kouch, P. M., Lioudakis, I., Fenu, F., et al. 2025b, *A&A*, 695, A99
- Kovalev, Y. Y., Petrov, L., Fomalont, E. B., & Gordon, D. 2007, *AJ*, 133, 1236
- Lähteenmäki, A., & Valtaoja, E. 2003, *ApJ*, 590, 95
- León-Tavares, J., Valtaoja, E., Tornikoski, M., Lähteenmäki, A., & Nieppola, E. 2011, *A&A*, 532, A146
- Lindfors, E. J., Hovatta, T., Nilsson, K., et al. 2016, *A&A*, 593, A98
- Lioudakis, I., Hovatta, T., Huppenkothen, D., et al. 2018a, *ApJ*, 866, 137
- Lioudakis, I., Romani, R. W., Filippenko, A. V., et al. 2018b, *MNRAS*, 480, 5517
- Lioudakis, I., Hovatta, T., Aller, M. F., et al. 2021, *A&A*, 654, A169
- Lioudakis, I., Hovatta, T., Pavlidou, V., et al. 2022a, *A&A*, 666, A36
- Lioudakis, I., Marscher, A. P., Agudo, I., et al. 2022b, *Nature*, 611, 677
- Lioudakis, I., Zhang, H., Boula, S., et al. 2025, *A&A*, 698, L19
- Lister, M. L., & Marscher, A. P. 1997, *ApJ*, 476, 572
- MAGIC Collaboration (Abe, S., et al.) 2024, *A&A*, 685, A117
- Malkin, Z. 2018, *ApJS*, 239, 20
- Marscher, A. P. 2014, *ApJ*, 780, 87
- Marshall, H. L., Lioudakis, I., Marscher, A. P., et al. 2024, *ApJ*, 972, 74
- Massaro, E., Maselli, A., Leto, C., et al. 2015, *Ap&SS*, 357, 75
- Mead, A. R. G., Ballard, K. R., Brand, P. W. J. L., et al. 1990, *A&AS*, 83, 183
- Meng, N., Zhang, X., Wu, J., Ma, J., & Zhou, X. 2018, *ApJS*, 237, 30
- Meyer, M., Scargle, J. D., & Blandford, R. D. 2019, *ApJ*, 877, 39
- Middei, R., Lioudakis, I., Perri, M., et al. 2023, *ApJ*, 942, L10
- Nalewajko, K. 2013, *MNRAS*, 430, 1324
- Nieppola, E., Hovatta, T., Tornikoski, M., et al. 2009, *AJ*, 137, 5022
- Nilsson, K., Pasanen, M., Takalo, L. O., et al. 2007, *A&A*, 475, 199
- Nilsson, K., Lindfors, E., Takalo, L. O., et al. 2018, *A&A*, 620, A185
- Otero-Santos, J., Acosta-Pulido, J. A., Becerra González, J., et al. 2022, *MNRAS*, 511, 5611
- Otero-Santos, J., Acosta-Pulido, J. A., Becerra González, J., et al. 2023, *MNRAS*, 523, 4504
- Paliya, V. S., Stalin, C. S., Ajello, M., & Kaur, A. 2017, *ApJ*, 844, 32
- Pandey, A., & Stalin, C. S. 2022, *A&A*, 668, A152
- Petrov, L., de Witt, A., Sadler, E. M., Phillips, C., & Horiuchi, S. 2019, *MNRAS*, 485, 88
- Planck Collaboration XVI. 2014, *A&A*, 571, A16
- Potter, W. J. 2018, *MNRAS*, 473, 4107
- Poutanen, J., Zdziarski, A. A., & Ibragimov, A. 2008, *MNRAS*, 389, 1427
- Raiteri, C. M., Villata, M., Acosta-Pulido, J. A., et al. 2017, *Nature*, 552, 374
- Ramakrishnan, V., Hovatta, T., Nieppola, E., et al. 2015, *MNRAS*, 452, 1280
- Rani, B., Gupta, A. C., Strigachev, A., et al. 2010, *MNRAS*, 404, 1992
- Readhead, A. C. S., Cohen, M. H., Pearson, T. J., & Wilkinson, P. N. 1978, *Nature*, 276, 768
- Rest, S., Rest, A., Wang, Q., et al. 2023, ATClean: High-Fidelity, Statistically Clean ATLAS Light Curves and Feature Detection
- Richards, G. T., Strauss, M. A., Fan, X., et al. 2006, *AJ*, 131, 2766
- Richards, J. L., Max-Moerbeck, W., Pavlidou, V., et al. 2011, *ApJS*, 194, 29
- Richards, J. L., Hovatta, T., Max-Moerbeck, W., et al. 2014, *MNRAS*, 438, 3058
- Romero, G. E., Cellone, S. A., Combi, J. A., & Andruchow, I. 2002, *A&A*, 390, 431
- Roy, N., Chatterjee, R., Joshi, M., & Ghosh, A. 2019, *MNRAS*, 482, 743
- Ruan, J. J., Anderson, S. F., MacLeod, C. L., et al. 2012, *ApJ*, 760, 51
- Safna, P. Z., Stalin, C. S., Rakshit, S., & Mathew, B. 2020, *MNRAS*, 498, 3578
- Sandrinelli, A., Covino, S., & Treves, A. 2014, *A&A*, 562, A79
- Scargle, J. D., Norris, J. P., Jackson, B., & Chiang, J. 2013, *ApJ*, 764, 167
- Schleicher, B., Arbet-Engels, A., Baack, D., et al. 2019, *Galaxies*, 7, 62
- Shin, M.-S., Sekora, M., & Byun, Y.-I. 2009, *MNRAS*, 400, 1897
- Sillanpää, A., Haarala, S., Valtonen, M. J., Sundelius, B., & Byrd, G. G. 1988, *ApJ*, 325, 628
- Smith, K. W., Smartt, S. J., Young, D. R., et al. 2020, *PASP*, 132, 085002
- Sokolovsky, K. V., Gavras, P., Karamelas, A., et al. 2017, *MNRAS*, 464, 274
- Sun, M., Xue, Y., Brandt, W. N., et al. 2020, *ApJ*, 891, 178
- Tachibana, Y., & Miller, A. A. 2018, *PASP*, 130, 128001
- Tonry, J. L., Stubbs, C. W., Lykke, K. R., et al. 2012, *ApJ*, 750, 99
- Tonry, J. L., Denneau, L., Heinze, A. N., et al. 2018, *PASP*, 130, 064505
- Valtaoja, E., Lähteenmäki, A., Teräsraanta, H., & Lainela, M. 1999, *ApJS*, 120, 95
- Vaughan, S., Edelson, R., Warwick, R. S., & Uttley, P. 2003, *MNRAS*, 345, 1271
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *Nat. Methods*, 17, 261
- von Neumann, J. 1941, *Annal. Math. Stat.*, 12, 367
- Wagner, S. M., Burd, P., Dorner, D., et al. 2022, in *37th International Cosmic Ray Conference*, 868
- Weaver, Z. R., Williamson, K. E., Jorstad, S. G., et al. 2020, *ApJ*, 900, 137
- Weisskopf, M. C., Brinkman, B., Canizares, C., et al. 2002, *PASP*, 114, 1
- Xiong, D., Sun, M., Wang, J.-X., et al. 2025, *ApJS*, 279, 3
- Zhang, B.-K., Zhou, X.-S., Zhao, X.-Y., & Dai, B.-Z. 2015, *Res. Astron. Astrophys.*, 15, 1784
- Zhang, H., Li, H., Guo, F., & Taylor, G. 2017, *ApJ*, 835, 125
- Zhang, T., Doi, M., Kokubo, M., et al. 2024, *ApJ*, 968, 71
- Żywucka, N., Goyal, A., Jamroz, M., et al. 2018, *ApJ*, 867, 131

## Appendix A: Variability measures other than $F_{\text{var}}$

Here we describe three metrics other than  $F_{\text{var}}$ , namely, IQR,  $1/\eta$ , and  $E_x$ , which we consider in this study for quantifying the overall variability of the CAZ light curves. In Sect. 4.1 we describe why we opt to focus on  $F_{\text{var}}$  as the main metric. Below we show the variability versus flux density plots for IQR,  $1/\eta$ , and  $E_x$  in Figs. A.1, A.2, and A.3, respectively.

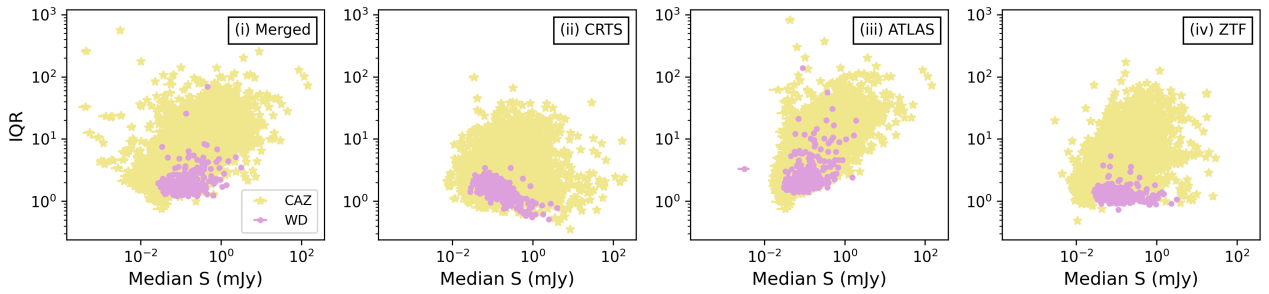
IQR is a scatter-based variability metric, equal to the difference between the 75<sup>th</sup> and 25<sup>th</sup> percentiles of the flux density distribution. We calculate it via  $\text{IQR} = (S_{75\%} - S_{25\%})/\text{MED}(\sigma_i)$ , where  $\text{MED}(\sigma_i)$  is the median flux density error and acts as a normalization factor. IQR is highly effective in distinguishing between variable and non-variable stars (Sokolovsky et al. 2017).

The von Neumann (1941) ratio,  $\eta$ , measures the mean square successive difference divided by the variance. Since variance increases with variability, a larger inverse ( $1/\eta$ ) indicates a larger degree of variability (e.g., Shin et al. 2009). This is calculated via  $1/\eta = \sum_{i=1}^N (S_i - \bar{S})^2 / \sum_{i=1}^{N-1} (S_{i+1} - S_i)^2$ , where  $\bar{S}$  is the sample mean. Notably,  $1/\eta$  is a correlation-based quantity which measures variability while taking the temporal order into account.

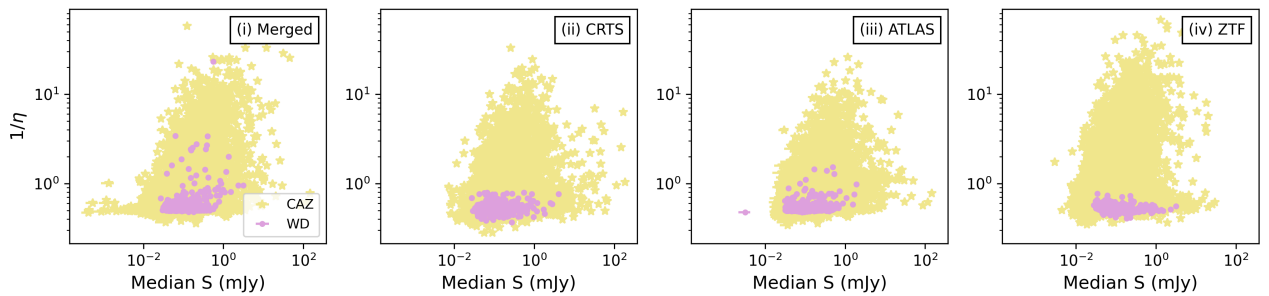
$E_x$ , another correlation-based variability metric, considers the observational gaps of a light curve (Sokolovsky et al. 2017). We divide the CAZ light curves into observational seasons by scanning for gaps longer than 60 d. Thus,  $E_x$  is calculated as:

$$E_x = \frac{2}{N_s(N_s - 1)} \sum_{i=1}^{N_s-1} \sum_{j>i}^{N_s} \frac{|\text{MED}(S_{s,i}) - \text{MED}(S_{s,j})|}{\sqrt{\text{MAD}(S_{s,i})^2 + \text{MAD}(S_{s,j})^2}} \quad (\text{A.1})$$

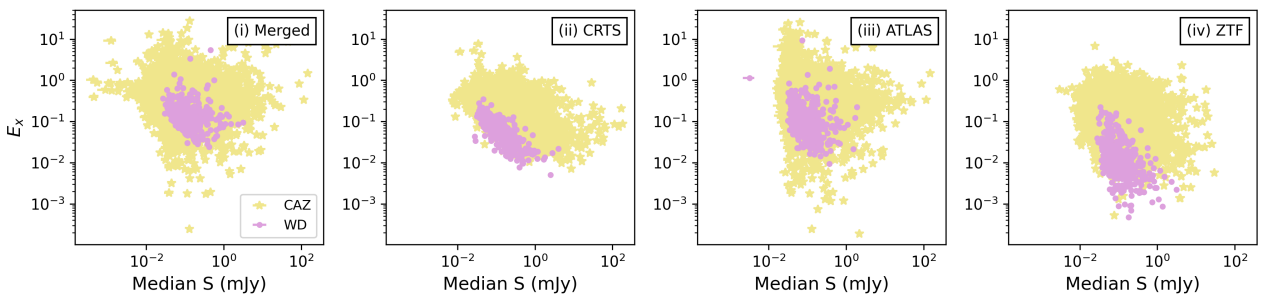
where  $N_s$  refers to the number of seasons and MAD is the median absolute deviation, calculated via  $\text{MAD}(S_{s,i}) = \text{MED}(|S_{s,i} - \text{MED}(S_{s,i})|)$  where  $S_{s,i}$  denotes flux densities within season  $i$  (likewise for season  $j$ ).



**Fig. A.1.** IQR versus median flux density plots. Plots (i–iv) show the scatter for merged light curves, CRTS-, ATLAS-, and ZTF-only, respectively.



**Fig. A.2.**  $1/\eta$  versus median flux density plots. Plots (i–iv) show the scatter for merged light curves, CRTS-, ATLAS-, and ZTF-only, respectively.



**Fig. A.3.**  $E_x$  versus median flux density plots. Plots (i–iv) show the scatter for merged light curves, CRTS-, ATLAS-, and ZTF-only, respectively.

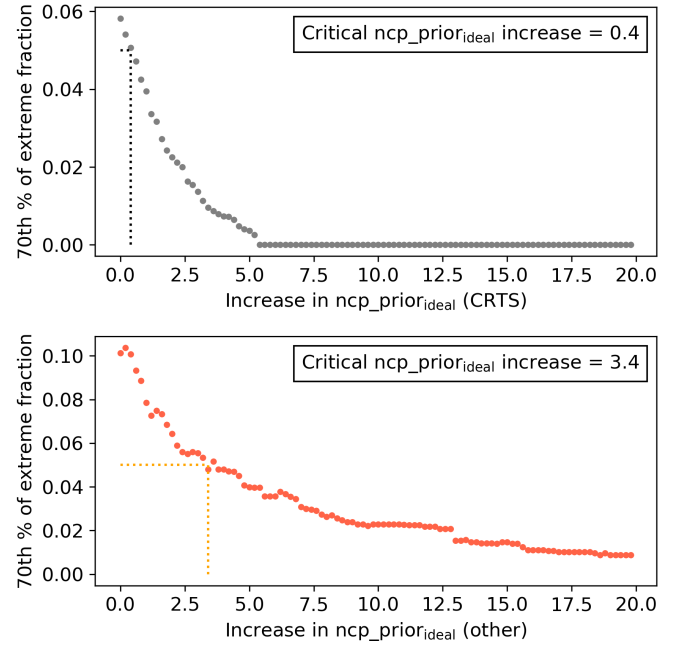
## Appendix B: Calibrating Bayesian blocks

As mentioned in Sect. 4.3, CAZ light curves of 408 WDs are used to calibrate `ncp_prior` of the BB algorithm in order to minimize the effect of noise. This is done by incrementally increasing `ncp_priorideal = 1.32 + 0.577 log10(N)` until the resulting BB light curve (with a total of  $N$  flux densities) is mostly consistent with a constant flux density corresponding to the median of the merged light curve. We note that CRTS and other (i.e., non-CRTS) filters are treated separately in this procedure.

We define an extreme WD BB fraction as the length of all BB whose flux density does not reach the median of the merged light curve within  $2\sigma$ , divided by the total length of all BB. In case of ideal Gaussian errors, we would expect this extreme fraction to be  $\sim 5\%$  because  $2\sigma$  Gaussian error bars should signify a  $\sim 95\%$ -likelihood range for the constant WD flux densities. However, in the case of WD CAZ light curves, `ncp_priorideal` generally results in extreme fractions greater than 5%. This demonstrates the noisy nature of the CAZ light curves, as mentioned multiple times throughout this paper.

To quantify the optimal value of `ncp_prior`, we incrementally add 0.2 to its ideal value (only dependent on  $N$ ) for each 408 WD light curves. We do this 100 times, covering a range of `ncp_prior = ncp_priorideal` to `ncp_prior = ncp_priorideal + 20`. For each step, we calculate the value within which  $\sim 1\sigma$  (70%) of the 408 extreme WD BB fractions are contained. This is plotted versus the incremental increase in `ncp_priorideal` for both CRTS and other filters separately in Fig. B.1. The calibrated increase in `ncp_priorideal` is defined as the value that results in  $\sim 1\sigma$  of the extreme fractions to be 0.05. This is 0.4 for CRTS, and 3.4 for other (non-CRTS) filters.

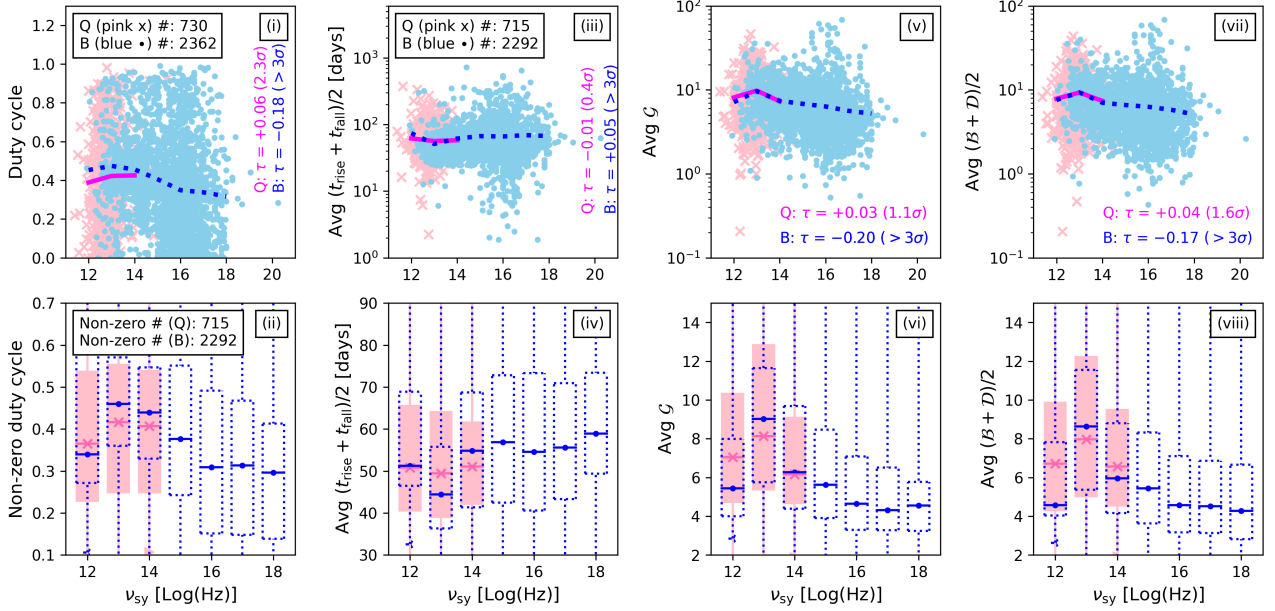
This result suggests that CRTS light curves are less noisy than the other filter light curves, which we confirm by visual inspection. ATLAS and ZTF light curves are especially prone to extreme outliers, noise-like behavior, and underestimated errors. This is why we select the global marker only at the  $\sim 1\sigma$  level, instead of a stricter one. For example, even as `ncp_prior`  $\rightarrow \infty$ , the extreme WD BB fraction of the non-CRTS light curves at the  $2\sigma$  level never reaches 0.05 (see the bottom plot in Fig. B.1).



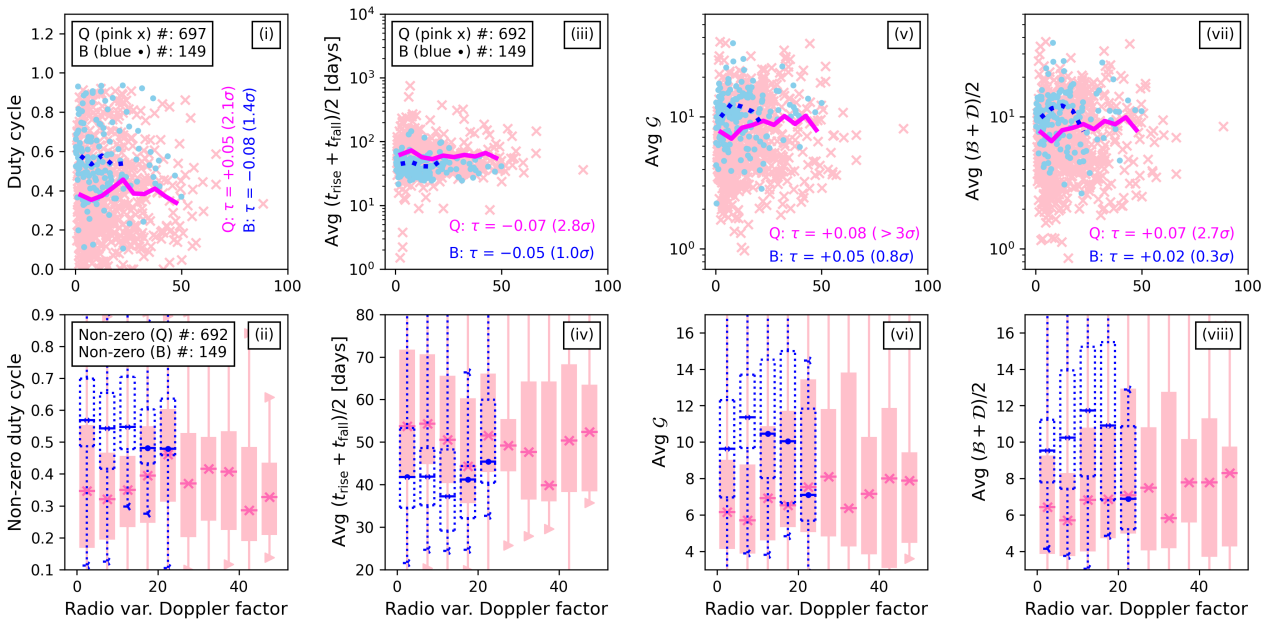
**Fig. B.1.** Value containing  $\sim 1\sigma$  (70%) of the extreme WD BB fractions at each incremental increase of `ncp_priorideal`. *Top:* CRTS WD light curves; *bottom:* other (i.e., non-CRTS) WD CAZ light curves. The dotted lines show the critical value of the increase in `ncp_priorideal` at the 0.05 level. It is 0.4 and 3.4 for CRTS and other filters, respectively.

## Appendix C: Characteristics of all BBHOP flares against physical parameters

As discussed previously (see Sect. 4.3 and 5.3), there are many noisy BBHOP flares. Thus, when analyzing their characteristics against synchrotron peak frequency and radio variability Doppler factor in Sect. 5.3, we only focused on a subset of prominent BBHOP flares (see Figs. 15 and 16). In this appendix we provide the same distributions for all BBHOP flares (not just prominent ones).



**Fig. C.1.** Distribution of source duty cycle (i and ii) as well as source-averaged average rise and fall times (iii and iv), global amplitude (v and vi), and average local brightening and decaying amplitudes (vii and viii) of all BBHOP flares against synchrotron peak frequency. Pink crosses and solid markings refer to FSRQs (Q); blue dots and dotted markings to BLLs (B). Kendall's  $\tau$  correlation coefficient and significance are reported for FSRQs and BLLs. In plot (i) we show zero duty cycle data, whereas in the other plots we do not. The running averages in the top row (shown as solid lines for FSRQs and dotted lines for BLLs) and the box plots in bottom row are both only calculated for at least ten nonzero values within each bin. In the top panels we show all data points, while in the bottom panels we zoom in on the boxes.



**Fig. C.2.** Similar as Fig. C.1 but for radio variability Doppler factor.

## Appendix D: Redshift-corrected flare parameters against radio variability Doppler factor

In Sect. 5.3.3 Fig. 16 we investigate the behavior of duty cycle as well as source-averaged average rise and fall times, and local and global amplitudes of prominent BBHOP flares against radio variability Doppler factor. While rise and fall times seem to decrease and amplitudes increase with increasing Doppler factor, here we investigate how correcting the flare parameters for redshift and luminosity distance changes the trends. We note that duty cycle (i.e., frequency of flaring) is not affected by redshift and, as such, is omitted from the investigation of this appendix.

As given in Kembhavi & Narlikar (1999), the intrinsic flare times are calculated via:

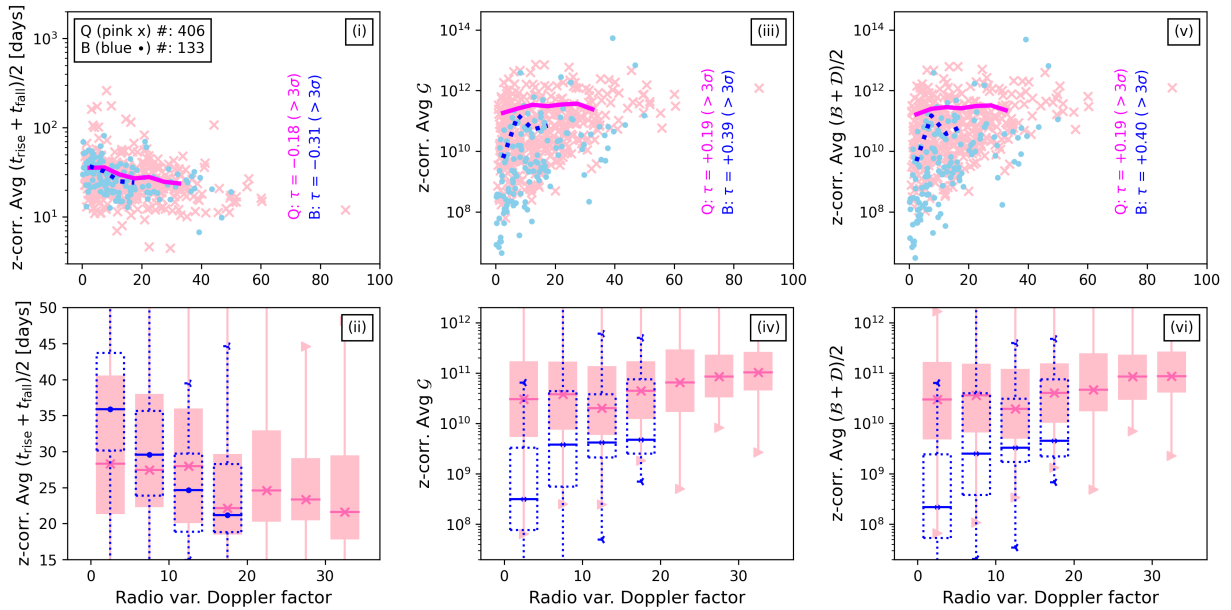
$$\Delta t_{\text{int}} = \Delta t_{\text{obs}} \cdot \left( \frac{\delta}{1+z} \right) \quad (\text{D.1})$$

where  $\Delta t_{\text{int}}$ ,  $\Delta t_{\text{obs}}$ ,  $\delta$ , and  $z$  are the source-frame flaring rise or fall time, the observed flaring rise or fall time, the Doppler factor, and the redshift, respectively. Meanwhile, the intrinsic flaring amplitude (in terms of luminosity) are calculated via:

$$\Delta L_{\text{int}} = \Delta S_{\text{obs}} \cdot \left( \frac{4\pi d_L^2}{1+z} \right) \cdot \left( \frac{1+z}{\delta} \right)^{3-\alpha} \quad (\text{D.2})$$

where  $\Delta L_{\text{int}}$ ,  $\Delta S_{\text{obs}}$ ,  $d_L$ , and  $\alpha$  are the source-frame flaring amplitude in terms of luminosity, the observed flaring amplitude in terms of flux density, luminosity distance, and spectral index as defined by  $S \propto \nu^\alpha$ , respectively. Spectral index in the optical band is class dependent. Here we use LSPs have  $\alpha = -1.1$ , ISPs  $\alpha = -1.3$ , and HSPs  $\alpha = -1.5$  (Fiorucci et al. 2004; Hovatta et al. 2014).

Since we correct the flare parameters with respect to redshift and luminosity distance, and then compare them to radio variability Doppler factor estimates, in Eqns. D.1 and D.2, we only focus on  $z$  and  $d_L$  while ignoring  $\delta$ . We divide the source-averaged average rise and fall times of the prominent BBHOP flares by  $1+z$ , and multiply their source averaged global and local amplitudes (which are proportional to the global and local  $\Delta S_{\text{obs}}$ , respectively) by  $4\pi d_L^2/(1+z)^{2-\alpha}$ . These redshift-corrected (“z-corr.”) flare parameters against radio variability Doppler factor are plotted in Fig. D.1. We note that to calculate  $d_L$  from  $z$ , we use the cosmology subpackage of *astropy* by assuming the cosmological constants of Planck Collaboration XVI (2014).



**Fig. D.1.** Distribution of redshift-corrected, source-averaged average flaring times (i and ii), global amplitudes (iii and iv), and average local amplitudes (v and vi) of prominent BBHOP flares against radio variability Doppler factor. For more plot details, see Fig. 16 where we show the same trends prior to redshift-correction.

As evident from Fig. D.1, the redshift-corrected flaring times become shorter and amplitudes larger as the radio variability Doppler factor increases. The  $>3\sigma$  negative correlation of redshift-corrected flare times with respect to radio variability Doppler factor is  $\tau = -0.18$  for FSRQs and  $\tau = -0.31$  for BLLs. Meanwhile, the  $>3\sigma$  positive correlation of redshift-corrected flare amplitudes with respect to radio variability Doppler factor is  $\tau = +0.19$  for FSRQs and  $\tau \approx +0.40$  for BLLs.

Notably, in both cases, the trends are more envelope-like than linear. At lower radio variability Doppler factors, a wider range of flare parameters are possible: faster and larger-amplitude flares as well as slower and lower-amplitude ones. However, at higher radio variability Doppler factors, there are fewer slow and low-amplitude flares. While this may be physical, it should be noted that it is, at least partially, because of host-galaxy contamination. Generally, blazars with lower Doppler factors are dimmer with respect to their host-galaxy and thus are more strongly affected by host-galaxy contamination, which diminishes their variability and our ability to identify prominent flares. This is likely why the trends are stronger in the case of BLLs than FSRQs, because BLLs are more strongly affected by host-galaxy contamination than FSRQs.