ASTRO WÜRZBURG

Statistical properties of the high-energy gamma-ray flux variations of blazars

Sarah Maria Wagner

Master's thesis at the Faculty for Physics and Astronomy
of the Julius-Maximilians-Universität Würzburg
Supervisors: Sara Buson & Karl Mannheim

Würzburg, May 2020

Abstract

There are about one hundred billion galaxies in the universe. Blazars are a very specific type of these galaxies featuring highly relativistic, bipolar jets that are directed towards us. The most energetic, steady radiation observed in the universe is identified with such sources. However, the mechanisms producing this radiation and its characteristic variability, are still uncertain. In this thesis, the statistical properties of the high-energy gamma-ray flux variations of blazars are studied in order to put constrains on the ground laying physical processes.

The first four statistical moments of the log-flux distributions from 335 monthly binned *Fermi*-sources are analyzed, indicating log-normalcy of the monthly flux distributions. Besides FSRQs being more variable than BLLs, no significant difference between the blazar sub-classes is evident. The statistical moments of the flux distributions from eight of these blazars in daily binning yield similar results. Additionally, these distributions of daily fluxes are fitted with a log-normal and stationarity is investigated through invariance of the mean flux on the one hand and a visual approach in the histograms on the other hand. The claim that log-normalcy implies the driving physical process to be multiplicative, is challenged along the lines of Scargle (2020). Multiplicative processes such as variations in the accretion flow that are inherited by the jet could be a possible explanation for the gamma-ray variability, nevertheless.

Time domain modeling of the gamma-ray flux variations in the monthly and daily sample is performed with Bayesian block analysis. Building up on that, two objective methods to define flares are introduced. The key question is whether these flares show a tendency for time (a)symmetry. The baseline method characterizes flares by the flux overcoming a certain threshold (Meyer et al., 2019). The half-clap method extrapolates the rising and falling behavior of the variations. Applying both of these methods to the monthly and daily sample shows that each ratio of rise and decay time occurs to a comparable degree, when excluding single-block flares. This exclusion is necessary because truly asymmetric flares can result in a (trivially symmetric) single-block flare when the time binning is too wide. The occurrence of every kind of asymmetry suggests that neither on monthly nor daily timescales, there is only one process causing distinct events. Possible explanations for the shape of the observed variations could be overlap or a mixture of processes. Moreover, it is likely that daily timescales are not enough to resolve the true shape of the high-energy variations, which is illustrated along the results of previous studies. Several approaches to refine the applied methods and approaches to remaining questions are presented in the end and considered for future work.

Zusammenfassung

Im Universum gibt es hundert Milliarden Galaxien. Ein sehr spezieller Typ davon sind Blazare. Sie zeichnen sich durch hoch relativistische Ausströmungen von Materie (Jets) aus, die eng zur Sichtlinie verlaufen. Von Blazaren geht variable Strahlung der höchsten überhaupt beobachteten Energien (gamma-Strahlung) aus, aber es ist unklar wie diese entsteht. In dieser Arbeit werden die statistischen Eigenschaften der gamma-Variabilität von Blazaren untersucht, um mögliche physikalische Prozesse ein zu schränken.

Aus einem Sample mit 335 monatlich gebinnten Quellen des *Fermi*-Satteliten werden die ersten vier statistischen Momente der Flussmessungen berechnet. Diese deuten eine log-normale Verteilung der monatlichen Flüsse an. Abgesehen davon, dass FSRQs variabler sind als BLLs ist (basierend auf den statistischen Momenten) kein Unterschied zwischen den Blazarunterklassen erkennbar. Das gleiche gilt für acht dieser Blazare bei täglich gebinnten Flüssen. Log-Normalität der täglichen Flüsse wird anhand von Fits diskutiert und Stationarität wird einerseits durch Zeitinvarianz des Mittelwertes und andererseits druch farbkodierte Histogramme untersucht. Die Behauptung, dass log-normale Flussverteilungen auf mutiplikative Prozesse schließen lassen, wird anhand von Scargle (2020) widerlegt. Nichtsdestoweniger kommen multiplikative Prozesse für die Erklärung der beobachteten gamma-Strahlung in Frage.

Mit dem Bayesian Block Algorithmus wird eine Zeitreihenanalyse der beiden Samples durchgeführt. Darauf aufbauend werden zwei Herangehensweisen zur Definition von Flares eingeführt um die Zeit(a)symmetrie der Flussfluktuationen zu parametrisieren. Während ein Flare in der Baseline Methode durch das Überschreiten eines festen Flusswertes bestimmt ist, wird bei der Half-Clap Methode das Steigungsverhalten der Variation extrapoliert. Mit beiden Methoden wird sowohl für die monatlich als auch für die täglich gebinnten Quellen jede Art von (A-)Symmetrie zu einem ähnlichen Anteil gefunden, wenn man den Überschuss an "einzelblock" Flares ausschließt. Dies ist nötig, weil ein tatsächlich asymmetrischer Flare als (trivial symmetrischer) "einzelblock" Flare dargestellt wird, wenn das Zeitbinning zu groß ist. Da jede Art von (A-)Symmetrie vorkommt, kann weder für das monatliche noch für das täglich gebinnte Sample in Frage kommen, dass nur ein bestimmter Prozess für die Flares auf diesen Zeitskalen verantwortlich ist. Überlappung oder eine Kombination aus mehreren Prozessen bietet eine mögliche Erklärung. Zusätzlich wird gezeigt, dass die Variabilität der gamma-Strahlung von Blazaren auf kürzeren Zeitskalen als Tagen stattfinden kann. Verschiedene Möglichkeiten um die eingeführten Methodik zu verfeinern und Ansätze für offene Fragen werden zum Schluss erwähnt.

Contents

1. Observing blazars in the gamma-regime

The first chapter of this work provides an outline of active galactic nuclei (AGN) and blazars. Numerous open questions concerning these sources are raised. These particularly refer to the relativistic jets which are thought to produce the high-energy radiation ($E > 1\,\text{MeV}$, also called gamma-radiation) that is of interest here. It is not clear which physical process underlay this high-energy emission. This work aims to study the flux variability of blazars in order to put constraints on the underlying physical processes.

1.1. Active galactic nuclei

About 20% of all galaxies are said to be active, meaning that they have a very luminous core. These so-called AGN produce their luminosity through accretion processes which outshine the ambient stellar emission. This is powered by the super massive black hole (SMBH) in the center of such galaxies.

Out of these AGN, another 20% feature bipolar, highly energetic outflows of matter. These are called jets and emit radiation in the radio regime. Thus, AGN exhibiting such jets are often referred to as radio-loud[1]. The unification scheme (Urry & Padovani, 1995), depicted in Fig. 1.1[2], represents the currently accepted structure of a radio-loud AGN. The SMBH in the center is surrounded by an accretion disk. Clouds with different properties that are depicted as different shaded blobs in the sketch cause characteristic

---

[1] Kellermann et al. (1989) classifies AGN as radio-loud if the ratio of radio (5GHz) and optical (B-band) flux exceeds a value of ten.

[2] Figure is not to scale; typical parameters according to Urry & Padovani (1995):
accretion disc $\approx 1 - 30 \cdot 10^{14}\,\text{cm}$, broad line region $\approx 2 - 20 \cdot 10^{16}\,\text{cm}$, inner radius of torus $\approx 10^{17}\,\text{cm}$, narrow line region $\approx 10^{18} - 10^{20}\,\text{cm}$, radio jets $\approx 10^{17} - 10^{24}\,\text{cm}$
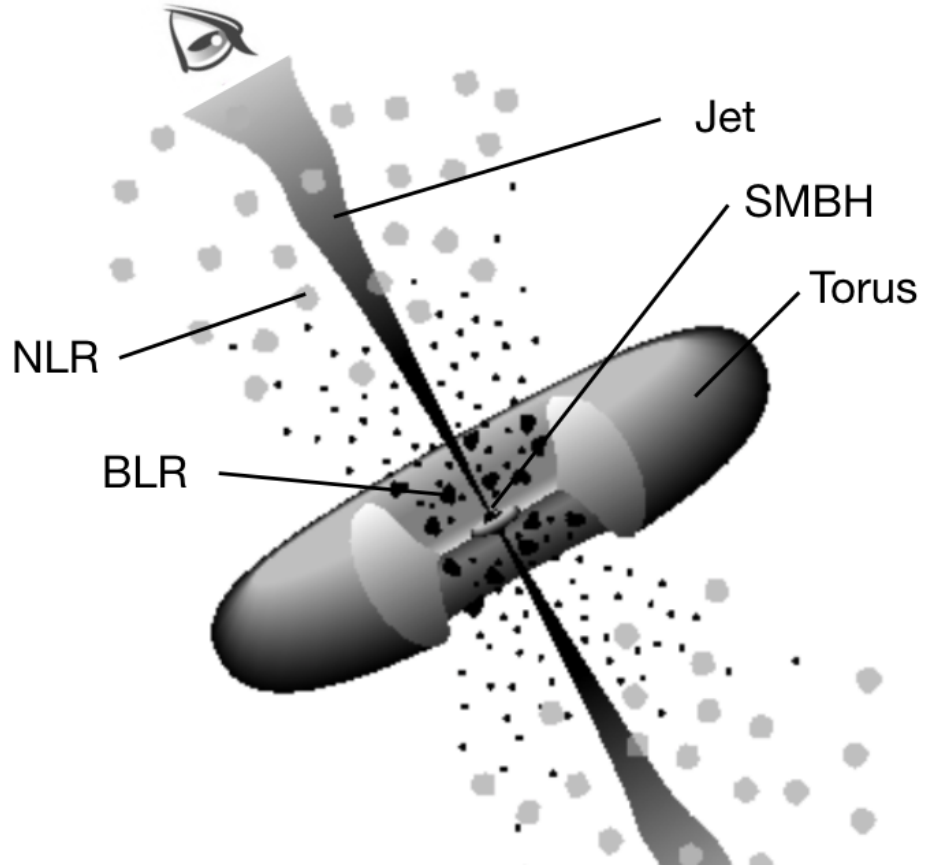
Figure 1.1.: Structure of an AGN with jets (not to scale; Urry & Padovani, 1995). Properties of the narrow line region (NLR), broad line region (BLR) and torus cause various emission features. Observation thereof is dependent on the angle of view. If the line of sight is close to the jet axis, as indicated with the eye, the AGN is called a blazar.

emission features. The big, grey blobs represent a low-density medium emitting narrow, forbidden emission lines. Hence, this is referred to as the narrow line region (NLR). Similarly, the so-called broad line region (BLR), which is depicted with black dots, produces Doppler broadened lines because the gas is closer to the black hole and therefore in rapid movement. Depending on the angle of view, these emission features can be obscured by the donut shaped dust torus surrounding the accretion disk.

For all sources considered in this work, the angle of view is closely aligned with the jet axis, as indicated with the eye. AGN observed with this angle of view are called blazars.

The bipolar jets launch in close vicinity to the black hole. "Despite [their] ubiquity,

many fundamental physics aspects of jets are still poorly understood and constrained," as summarized in a detailed overview on jets from AGN by Boettcher et al. (2012). One of these questions is as to why jets only occur in 20% of all AGN and how the launching mechanism works in general. The commonly consulted theories are that the jet extracts electromagnetic energy from the rotation energy of the SMBH (Blandford & Znajek, 1977) and/or that the jets are driven by accretion disc winds (Blandford & Payne, 1982). A further topic of recent research is how the jets stay collimated and transport energy, momentum, and angular momentum over their vast extension up to mega parsecs.

VLBI studies yield that the jets consist of components of higher flux density that are traveling outwards (e.g. Jorstad et al., 2017; Wagner, 2018). Due to the small angle of view, the speed of such components can appear to be faster than the speed of light. Another effect of this geometry is that the radiation is Doppler-boosted (Rybicki & Lightman, 1986). Thus, the emission from the jet dominates and often even outshines the host galaxy of blazars.

1.2. Blazars

The spectral energy distribution (SED) illustrates the behavior of flux density F at the corresponding frequency ν. The SED is typically plotted in logarithmic scale and the flux density is multiplied with frequency such that the integrated area under the graph corresponds to the received power of a given frequency range (per area). For a blazar, this results in a characteristic double humped structure as can be seen in the SED of PKS 1510-089 shown in Fig. 1.2.

Due to the fact that the radiation associated with the low-energy peak shows polarization, it is expected to be produced via synchrotron emission from an electron distribution.

The origin for the second hump in the SED is topic of recent research. Commonly, it is modeled with inverse Compton (IC) scattering. This means the relativistic electron distribution causing the synchrotron emission additionally up-scatters ambient photons to high energies. These ambient photons could be the synchrotron photons produced before which is referred to as synchrotron self Compton (SSC) model. Models based on
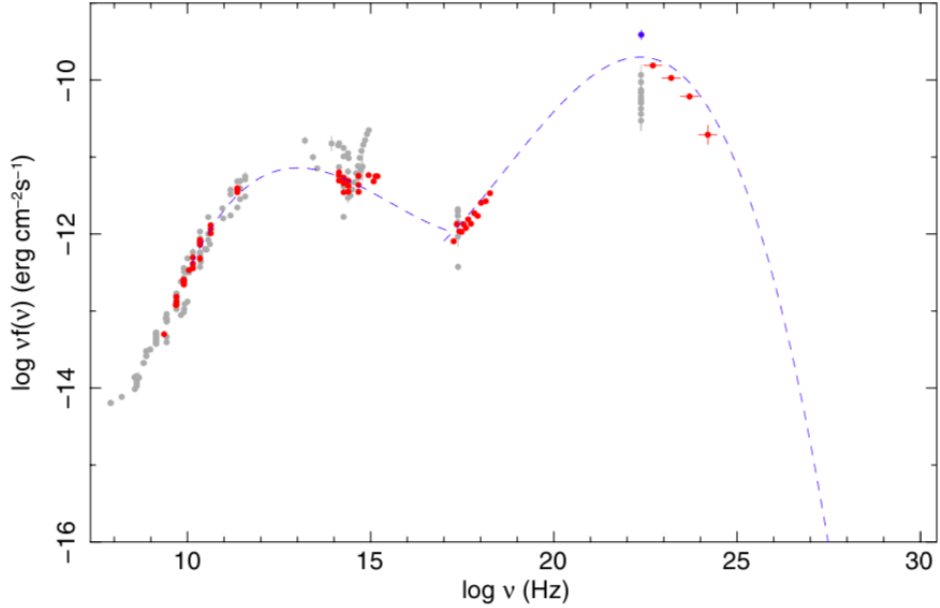
Figure 1.2.: The SED of PKS 1510-089 shows the characteristic double humped structure of a blazar. Red points show quasi simultaneous measurements that are fitted with a polynomial while the grey points represent archival data. Figure adapted form Abdo et al. (2010).

other photon sources are called external Compton (EC) models. The external photons could originate from the broad line region, the accretion disk, the dust torus, or even other components of the jet (see Finke, 2016, and references therein for an overview of IC models). In addition to the electrons, Mannheim (1993) first suggested that protons are also present in the jet. These could cause a contribution to the high-energy hump via synchrotron radiation or interaction with ambient photons and would also produce neutrinos.

Blazars can be sub-classified based on the position ν_{syn} of their synchrotron peak. Abdo et al. (2010) classify blazars as low-synchrotron peaked (LSP) if $\nu_{syn} < 10^{14}$ Hz, intermediate-synchrotron peaked (ISP) if 10^{14} Hz $< \nu_{syn} < 10^{15}$ Hz, and high-synchrotron peaked if $\nu_{syn} > 10^{15}$ Hz. It has to be noted that the synchrotron peak and therefore this classification can change over time.

Another blazar sub-classification is based on the presence of emission lines. Historically, blazars without emission lines were called BL Lac objects (BLLs). This tracks

back to BL Lacertae, the first source identified of that kind (Schmitt, 1968). Blazars that do show emission features are called Flat Spectrum Radio Quasars (FSRQs). It eventually became clear that the emission lines are sometimes more and sometimes less outshone by the broadband emission of the jet. Even the prototype BL Lacertae itself can show emission lines (Vermeulen et al., 1995). Thus, Ghisellini et al. (2011) express the distinction based on the ratio of broad line luminosity to Eddington luminosity being less than $5 \cdot 10^{-4}$ for BLLs and greater than that for FSRQs, respectively. This is physically motivated by the thought that the "presence of strong emitting lines is related to a transition in the accretion regime, becoming radiatively inefficient below a disc luminosity of the order of $1\,\%$ of the Eddington one". All blazars classified as HSP are BLLs and most LSP blazars are FSRQs.

1.3. Blazar variability

As can be seen in Fig. 1.2, the flux density at a given frequency is not constant. In fact, variability throughout the whole electromagnetic spectrum is typical for blazars (Urry, 1999). This study aims to analyze the high-energy variability behavior of blazars in order to put constraints on the processes underlying these variations.

In a first approach, the distributions of the observed fluxes are considered. Variability studies typically account for a fit of the analyzed flux distributions (e.g. Meyer et al., 2019; Tluczykont et al., 2010; Ackermann et al., 2015). They mainly discuss the distinction of normal and log-normal distributions (Shah et al., 2018). This is motivated by the hope of tracking properties of the distribution back to the process producing the flux. This process is often assumed to be composed of various sub-processes. The overall process is referred to as "additive" if it is composed by the sum of the sub processes or as "multiplicative" if it is based on the product of these.

Recent studies, such as Shah et al. (2018), point out that gamma-ray flux distributions show log-normal rather than normal behavior. Stating that this requires the ground laying process producing the gamma-ray emission to be multiplicative, Rieger (2019) poses possible explanations. One that is commonly evoked in other studies as well is that variations in the accretion flow of the disk propagate inwards and are inherited by the jet, where the gamma-rays are produced (see Sec. 3.1). This hypothesis is studied

by analyzing the log-normality of flux distributions from blazars. In addition to that, stationarity of the flux distributions is illustrated and discussed.

An independent, second approach to learn about emission processes present in blazars is based on time domain analysis. One can generally discriminate between physically motivated (physical) and data driven (non-physical) time domain modeling. The former evokes physical assumptions that could be based on constrains from e.g. acceleration, cooling, dynamical, and instability timescales. These can give implications on the rising and falling behavior of the flux which justifies to model variations, for instance, with an exponential fit. The goal of this work is to constrain the physical assumptions based on the observed variability behavior. Therefore, non-physical time domain modeling is chosen to objectively characterize the flux increases and decreases. Common non-physical methods to do so are Fourier Analysis, auto-correlation, or Bayesian block analysis (see Sec. 2.2). Based on the latter, two methods that define the rising and falling behavior of the flux are introduced. Of the many properties that can be obtained with these methods, the timescales of rise and fall are of concern for this work. Studies with similar goals often exclusively discuss single flux variations with high flux amplitude that are called flares (e.g. Hovatta et al., 2015, and many more). By contrast, this study aims to conduct a systematic analysis of all variations in a large sample of sources. For this thesis, a flare is defined by the corresponding method (see Sec. 2.2). With systematic comparison of detected flares in a large sample of light curves this thesis aims to provide constraints for physical models of the emission processes.

1.4. FERMI-LAT

Gamma-rays interact with the atmosphere through Compton scattering, the photoelectric effect, and pair production. The latter dominates at the higher energies ($> 1\,\mathrm{MeV}$) as are of interest here. If the initial photon energy is high enough the produced pair causes a leptonic cascade of photons and electrons in the atmosphere which in turn can cause Cherenkov light which is studied by ground-based telescopes such as CTA[3] or MAGIC (Lorenz & MAGIC Collaboration, 2004).

In order to observe gamma-rays of moderate energies, measurements have to be con-

---

[3] See https://arxiv.org/pdf/0906.4114.pdf, 11.02.2020

ducted on satellites above the atmosphere. Currently operating missions are the International Gamma-Ray Astrophysics Laboratory[4] (INTEGRAL), the Astro-Rivelatore Gamma a Immagini Leggero[5] (AGILE), and the *Fermi* Gamma-ray Space Telescope[6]. The latter is the most recent and most sensitive space-based gamma-ray telescope so far and is the source of data studied in this work.

The *Fermi*-sattelite was originally named Gamma-ray Large Area Space Telescope (GLAST) when it was launched on the 11th of June 2008. The NASA-mission device has been assembled at the SLAC National Accelerator Laboratory[7] (SLAC) and carries two instruments as scematically shown in Fig. 1.3. The Gamma-Ray Burst Monitor (GBM) is used to study gamma-ray bursts while the major instrument, the Large Area Telescope (LAT), conducts an all-sky survey for gamma-rays with energies from 20 MeV to greater than 300 GeV (Atwood et al., 2009). The large field of view (2.4 sr) covers one fifth of the sky which, combined with the 90 minute orbit around earth and a specific rocking motion, allows to monitore the entire sky every three hours. The *Fermi*-LAT contains four components:

1. The **Tracker** consists of several layers of silicon-strip particle tracking detectors. Therein, an incoming photon is converted to an electron and positron pair the paths of which are immediately measured. The presence of two coinciding paths allows the differentiation of gamma-ray photons and charged particles (cosmic rays).

2. The **Anticoincidence Detector** is used to filter cosmic rays. On average, there is one gamma-ray photon in ten thousand to one million background events. The instrument consists of specially formulated plastic tiles that produce flashes of light when hit by a cosmic ray.

3. The **Calorimeter** measures the energy the secondary electron-positron pair through a proportional intensity of a light flash caused by interaction with the material (cesium iodide).

4. The information captured by all this is interpreted with the microprocessors of the

---

[4] See http://www.esa.int/Enabling_Support/Operations/Integral_operations, 30.01.2020

[5] See http://www.agilescienceapp.it/wp/agilescienceen/?page_id=685, 30.01.2020

[6] See https://www.nasa.gov/content/fermi-gamma-ray-space-telescope, 30.01.2020

[7] See https://www6.slac.stanford.edu, 30.01.2020
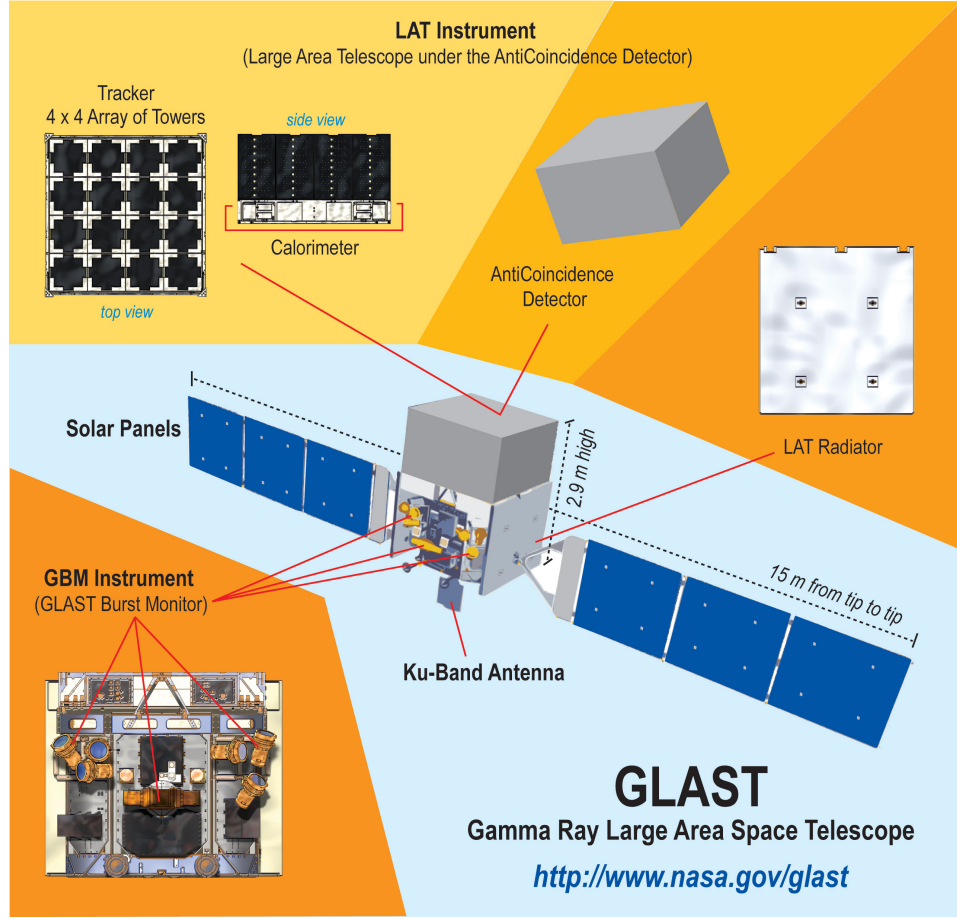
Figure 1.3.: Schematic set up of the *Fermi* Gamma-ray Space Telescope.

Data Aquisition System, such that only gamma-ray events are relayed to the ground.

The *Fermi*-LAT data is available as so-called Pass 8 data on the *Fermi*-website[8]. This data has been preprocessed regarding event reconstruction, energy reconstruction, and further classification such as identifying cosmic rays and multiple events as explained in Atwood et al. (2013). After choosing a certain location, search radius, and energy range, one can download the coresponding fits files containing information for further analysis (photon file with event measurements and space craft file with telescope properties).

---

[8] See https://fermi.gsfc.nasa.gov/cgi-bin/ssc/LAT/LATDataQuery.cgi, 30.01.2020

	Name	Class	Type	redshift
0	OJ 287	BLL	LSP	0.306
1	PKS 0208-512	BLL	LSP	1.003
2	PKS 0235+164	BLL	LSP	0.94
3	PKS 1510-089*	FSRQ	LSP	0.360
4	PKS 2155-304	BLL	HSP	0.116
5	3C 273*	FSRQ	LSP	0.158
6	3C 279*	FSRQ	LSP	0.536
7	3C 454.3*	FSRQ	LSP	0.859

Table 1.1.: Sources and parameters of the daily sample covering August 2008 up to January 2018 with energy cut at $E > 100\,\text{MeV}$, where * indicates publication in Meyer et al. (2019).

The *Fermi*-Lat collaboration provides recommendations for these steps[9] depending on the desired type of analysis for each scientific goal.

The main data set for this study was originaly employed in the chance coincidence analyses of the neutrino IC170922A and the blazar TXS 0506+056 (IceCube Collaboration et al., 2018). The data spans the time interval from August 2008 to October 2017 and is binned in time in 28 day intervalls. The sample, hereafter referred to as "monthly" sample, contains 2276 sources and covers energies $E > 1\,\text{GeV}$. It has to be noted that although not all of these are classified as blazars, they are all considered as such for the purpose of this work (see Sec. 2.1.1).

In addition to that, these findings are compared to a smaller sample of eight sources in daily binning, which will be referred to as "daily" sample. A list of it is shown in Tab. 1.1. Four of these sources were already discussed in the variability study Meyer et al. (2019), which is denoted with *. Additional information was obtained from the third catalog of AGN detected by *Fermi* (Ackermann et al., 2015, 3LAC[10]). This includes the object class (FSRQ or BLL)[11] determined with optical observations, the type (HSP, ISP, or LSP) based on the SED classification, and the redshift. The measurement time span

---

[9] See https://fermi.gsfc.nasa.gov/ssc/data/analysis/documentation/Cicerone/, 30.01.2020

[10] See https://www.ssdc.asi.it/fermi3lac/, 03.03.2020

[11] PKS 0235+164 is a blazar of unknown type according to 3LAC but classified as BLL on Simbad: http://simbad.u-strasbg.fr/simbad/sim-id?Ident=PKS+0208-512&submit=submit+id, 03.03.2020

Sample name	"monthly"	"daily"
Time binning	28 days	24 hours
Energy cut	$E > 1\,\mathrm{GeV}$	$E > 100\,\mathrm{MeV}$
Sources	2276	8
Time span	Aug 2008 - Oct 2017	Aug 2008 - Jan 2018
Published in	IceCube Collaboration et al. (2018)	Meyer et al. (2019)

Table 1.2.: Properties of the two data sets studied in this work.

covers August 2008 up to January 2018 and the energy cut was set to $E > 100\,\mathrm{MeV}$.

The properties of both samples are summarized in Tab. 1.2. Within this work the terms "flux value", "observed flux", "bin", or "data point" will always refer to the flux value derived for a constant time bin (monthly or daily, respectively).

Each flux value obtained in *Fermi* analysis is associated with a value called test statistic (TS). It is used to describe the certainty that an observed photon is in fact from the suspected source rather than from another one. It is defined as the log-likelihood ratio between the maximized likelihoods $\mathcal{L}_1$ and $\mathcal{L}_0$ for the hypothesis with and without an additional source, respectively (Mattox et al., 1996). Based on this, TS $= -2ln(\mathcal{L}_1/\mathcal{L}_0)$ can be interpreted as the squared detection significance such that TS > 4 equals a detection significance greater than two sigma. The most common, yet ambiguous, way to work with data form *Fermi* analysis is to choose a limit for significant detection and treat detections not fulfilling this condition as "upper limits" (e.g. Peñil et al., 2020). Nevertheless, upper limits can be problematic for further processing of the data. This work requires TS > 4 where all bins not fulfilling this condition will be neglected (see Sec. 2.1.1).

2. Data analysis and methodology

After the *Fermi*-analysis, there are several ways to further study the measured fluxes. Two are presented in this chapter. First, the general flux distribution is considered. This involves fitting the distribution in form of a histogram and calculating the statistical moments thereof. The ambiguity of the production and interpretation of these results is discussed in Sec. 3.1. The other approach utilizes non-physical time domain modeling in form of Bayesian blocks. Building up on these, two systematic methods to define flares and thereby characterize the flux variability behavior are introduced. All of this analysis is performed with python version 3.6.8 (Van Rossum & Drake, 2009) and made use of Astropy[1], (see Appendix).

2.1. Flux distributions

When consulting histograms, it is not immediately clear which binning to use. The choice of binning, however, can affect the representation of the data by suppression or destruction of important features. Often the data range is divided into an arbitrary, equally sized number of bins. A more sophisticated way to choose the number of bins is introduced by Knuth (2006)[2]. The optimal number of equally sized bins is determined by "assigning a multinomial likelihood and a non-informative prior [in order to] derive the posterior probability for the number of bins in a piecewise-constant density model given the data." Another approach is to use Bayesian blocks according to Scargle et al. (2013). In this case, the binning structure is adapted to the data which conveniently brings out all significant features. Due to the fact that the Bayesian bins can vary in

---

[1] Astropy is a community-developed core Python package for Astronomy http://www.astropy.org, 26.05.2020

[2] Utilized implementation: https://docs.astropy.org/en/stable/api/astropy.visualization.hist.html, 04.02.2020

size, wide bins will naturally contain more data points than narrow ones. As a constant size of bins is easier accessible, the binning of all histograms plotted in this work is determined by the Knuth algorithm.

2.1.1. Sample study of the monthly sources

Properties of the monthly flux distributions are compared based on the source classes (FSRQ, BLL, and unclassified). To allow comparison between individual sources, **all flux bins** F_i (typically of the order $10^{-8} - 10^{-6}\,\mathrm{GeV\,cm^{-2}\,s^{-1}}$) **are scaled by the mean** $\overline{F}$ of the corresponding light curve. Thus, the flux values are mainly distributed around zero when **plotted in logarithmic scale**. This is shown in the top left of Fig. 2.1 for the distribution of all scaled fluxes as obtained from the analysis explained Sect. 1.4. Besides the expected distribution of fluxes around zero, a distinct distribution of smaller flux values shows up as well. This illustrates that several flux values are significantly lower than the corresponding mean flux of their source. The distinct distribution vanishes if only flux values with TS > 4 (see Sect. 1.4) are included, such as shown in the top right of Fig. 2.1. Thereby, the distinct distribution is identified as noise/background. Individual flux bins with TS < 4 are consequently neglected from here on. While in the literature these are sometimes considered as "upper limits" (see Peñil et al., 2020), they will be handled as gaps here. Consequently, the mean $\overline{F}$ (used as scaling factor) is calculated by the corresponding TS > 4 flux values only.

One quarter of all flux bins in the monthly sample fulfill the condition TS > 4. In order to avoid a bias towards higher fluxes, sources with more than 50% of the flux bins missing are excluded from this analysis (analogous to Peñil et al., 2020). This affects 1941 sources, while the remaining 335 sources are considered for further analysis. The resulting distribution of fluxes is shown in the bottom of Fig. 2.1.

Lastly, it has to be noted that the original sample of 2276 sources contains several classes of extra-galactic objects as anticipated for its initial purpose. This involves 416 FSRQ and 618 BLL associations. In addition, 37 FSRQs and 18 BLLs are identified with counterpart. The filtered sample of 335 sources used here contains 69 (12) FSRQs and 106 (2) BLL associations (additional identifications). The remaining sources are unclassified (UC) AGN. This is due to the fact that these classifications require observations in other wavelengths (see previous section) that are not available for the large
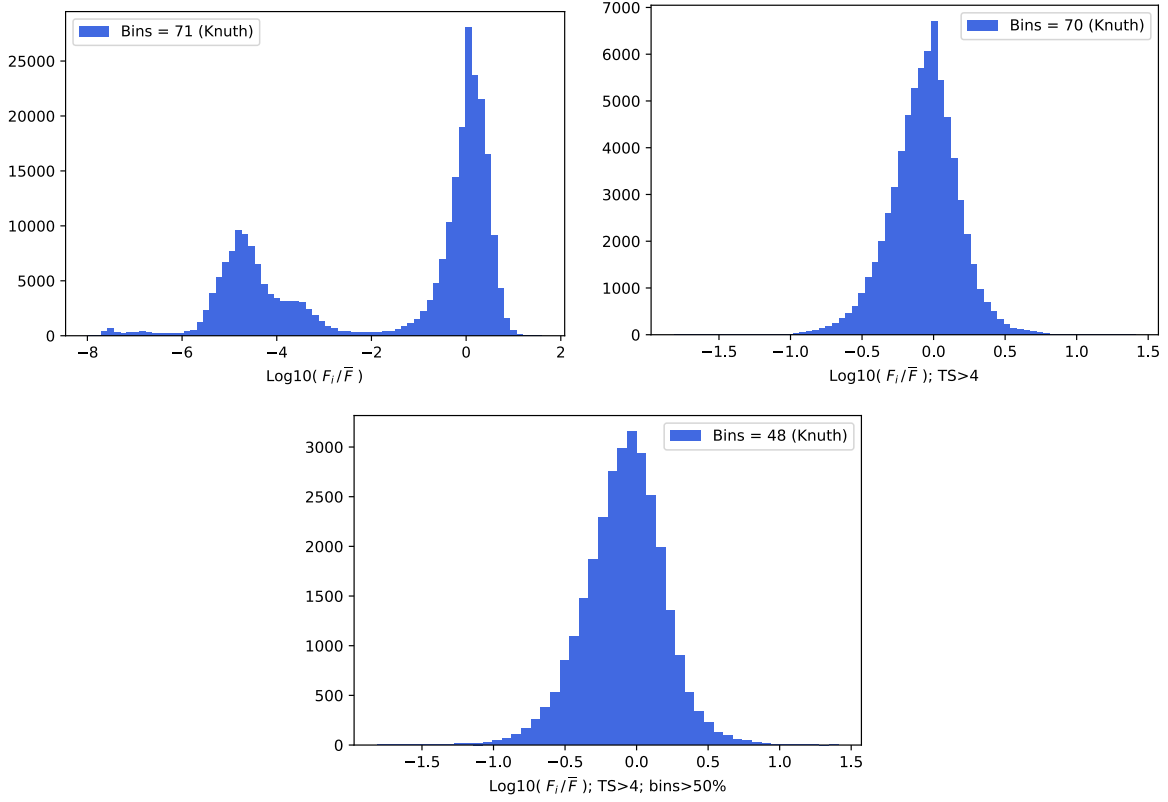
Figure 2.1.: Distribution of every individual flux bin of the monthly sample scaled by the corresponding mean of each light curve in logarithmic scale. *Top left:* initial sample. *Top right:* only bins with TS > 4. *Bottom:* contains only sources where at least 50% of all bins have TS > 4. This filter is used throughout this work.

$$
\begin{aligned}
\text{1st}: \quad & \frac{1}{n}\sum_i^n x_i = \mu && \text{Mean} \\[2ex]
\text{2nd}: \quad & \frac{1}{n}\sum_i^n (x_i - \mu)^2 = \sigma && \text{Variance} \\[2ex]
\text{3rd}: \quad & \frac{1}{n}\sum_i^n \frac{(x_i - \mu)^3}{\sigma^3} && \text{Skewness} \\[2ex]
\text{4th}: \quad & \frac{1}{n}\sum_i^n \frac{(x_i - \mu)^4}{\sigma^4} && \text{Kurtosis}
\end{aligned}
\tag{2.1.1}
$$

amount of sources detected by FERMI. Nevertheless, the unclassified sources have similar properties to the ones classified as balzars, for instance their spectra. This can be shown by applying artificial neural networks, as performed in Liodakis & Blinov (2019) and Kovačević et al. (2020) or it is visible when comparing the power density spectra of all sources in the sample (P.R. Burd, private conversation). Therefore, it is assumed that all unclassified sources of this sample are blazars as well.

To sum up, the initial sample of 2276 sources reduces to 335 by requiring TS > 4 for each flux bin and a minimum of 50% remaining bins for each source. All of these sources are assumed to be blazars.

To characterize the log-flux distribution of each source, the first four statistical moments are computed. For n random variables x_i, these are defined as shown in Eq. 2.1.1. In this work, the random variable x_i equals the logarithmic, scaled flux values: $\log(F_i/\overline{F})$. For real measurements, sampling adjustments accounting for the degrees of freedom have to be added. This is performed for the individual first, second, third[3], and fourth[4] statistical moment of the log-flux distribution from each source. The normalized distributions of the resulting parameters are shown in Fig. 2.2 for the source class FSRQ (red shaded), BLL (blue dotted), and unclassified (grey contour).

The first statistical moment is the mean μ (first panel of Fig. 2.2) of the logarithmic fluxes observed for each individual source. The flux is dependent on the distance of the

---

[3] Utilized implementation: https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.skew.html, 04.03.2020

[4] Utilized implementation: https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.kurtosis.html, 04.03.2020
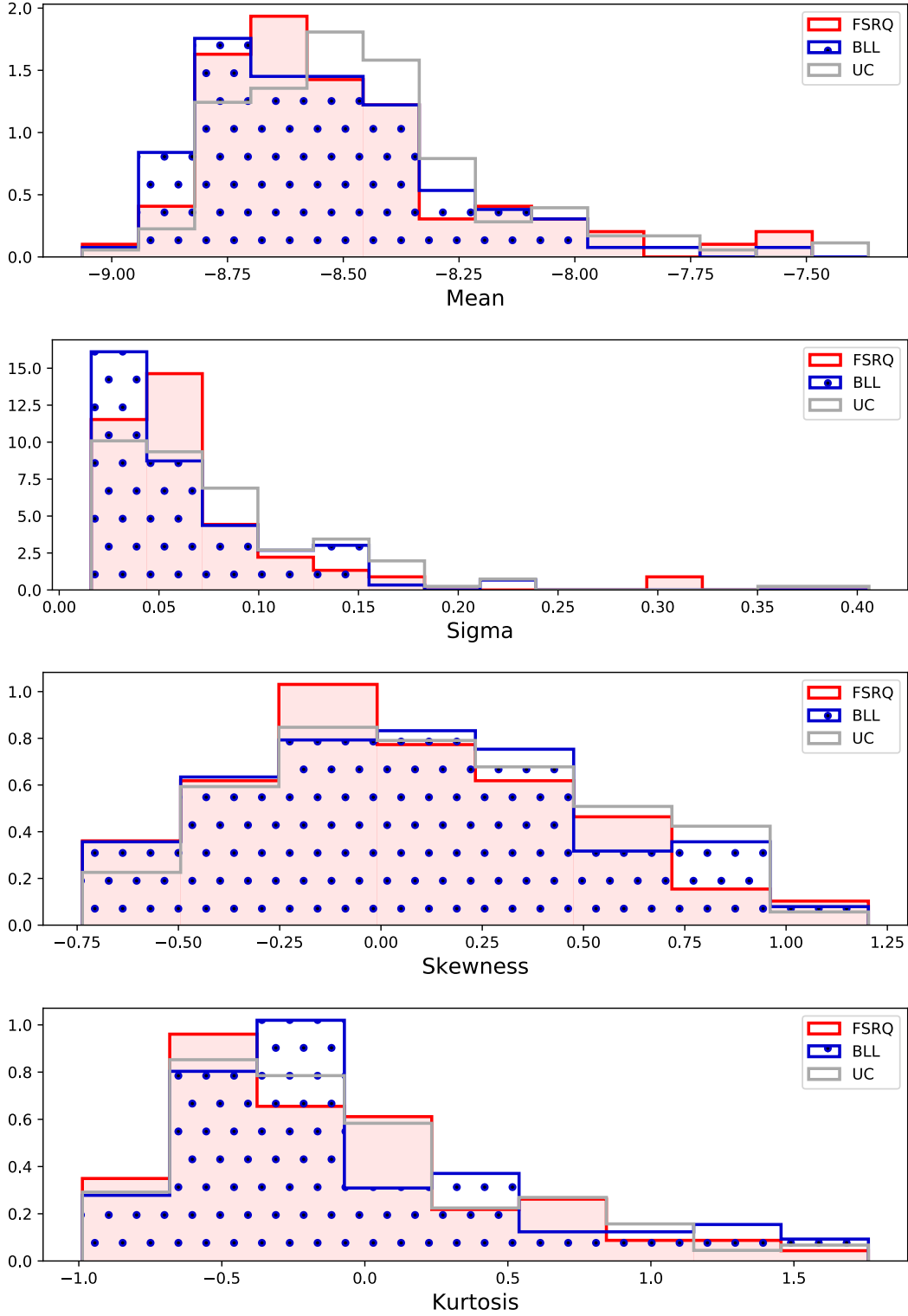
Figure 2.2.: Normalized distributions of first, second, third, and fourth (top to bottom) statistical moment of the log-flux distribution of each individual source of the monthly sample according to Eq. 2.1.1. The resulting values are shaded red, dotted blue, and contoured grey according to source class "FSRQ", "BLL", and "UC".

source[5]. Hence, these mean values are unsuitable for comparison of class properties.

The second moment, variance or σ (second panel of Fig. 2.2), can be interpreted as the range of variability (Ackermann et al., 2015). It is independent of the concrete value that the flux it is varying around. Thus, comparison between the source classes is possible. It can be seen in the first two histogram bins of Fig. 2.2 that most FSRQs show larger variances than most BLLs. This indicates that FSRQs are more variable than BLLs, which is in accordance with previous results reported in Ackermann et al. (2015). Nevertheless, every source class shows instances of high variability. The most variable sources ($\sigma > 0.25$) are 3FGL J1224.9+2122, 3FGL J1504.4+1029, 3FGL J2232.5+1143, and 3FGL J2254.0+1608 (3C 454.3). This histogram fits the assumption that the unclassified sources are blazars and therefore constituted of a mixture of FSRQs and BLLs.

The skewness, or third statistical moment (third panel of Fig. 2.2), indicates asymmetry of a given distribution. While zero represents symmetry, values more than ±0.5 (±1) away from zero are usually considered to be moderately (highly) asymmetric. Negative values indicate a left skew, which means that the distribution shows a long tail to the left with an accumulation of values on the right. Positive skewness values indicate a tail to the right and a bulk on the left accordingly. The skewness values for the monthly log-flux distributions range from -0.75 to 1.25, as shown in the third plot of Fig. 2.2. Most sources are approximately symmetric. The skewness has to be treated with caution in these cases because the data included here might lack small values that are below the sensitivity threshold of the satellite or were excluded due to TS< 4. The overall distribution of skewness values is the same for FSRQs, BLLs, and UCs, which is in agreement with UCs consisting of FSRQs and BLLs.

The bottom plot of Fig. 2.2 shows the distribution of the fourth statistical moment, which is also called kurtosis. Historically, the kurtosis was also referred to as the "peakedness" of a distribution. By now it is established that it rather describes the "thickness of the tails" of the distribution[6]. In this implementation, a kurtosis of zero corresponds to a normal distribution. A distribution that is more shaped like a box (i.e. more values close to the mean, also referred to as "stocky" or "broad shoulders") would have a negative kurtosis up to -4, while a distribution with long, thick tails has a positive kurtosis

---

[5] Assuming that the luminosity is constant, the flux is only dependent on distance. This will be used for the daily sample to compare individual sources.

[6] See https://doi.org/10.1080/00031305.2014.917055, 11.05.2020

up to infinity. The kurtosis parameters of this sample range from -1 to 1.5 with BLLs showing slightly larger values than the FSRQs.

A two sample Kolmogorov-Smirnov test[7] is utilized to quantify the hypothesis that the distribution of parameters pictured in Fig. 2.2 are drawn from the same sample. This hypothesis can be rejected for the variance with a p-value of 5%. For all other statistical moments it can not be ruled out that the distributions of each class belong to the same overall distribution. To conclude, the FSRQs of the monthly sample are more variable than the BLLs. Considering the first to fourth statistical moment, no other differentiation between the classes is possible.

2.1.2. Individual flux distributions of the daily sample

The flux bins of the daily sample are also filtered for TS > 4. The resulting flux distribution of each source is shown by introducing color coded histograms. Additionally, the distributions are fitted with a normal in log scale, which corresponds to a log-normal fit.

Presenting a sequence (e.g. time, frequency, energy band) of data in a histogram does not conserve the information which possibly is encoded within the order of the sequence. The possibility of representing this information visually is explored here by plotting the histogram in a color scheme according to time. This is demonstrated in in Fig. 2.3 - 2.10 for the daily sample.

The top plot shows the distribution of flux values scaled to the mean of each light curve. It is plotted in logarithmic scale such that the values are again distributed around zero. The plot underneath depicts the light curve of each source with the time spanning along the x-axis and the scaled flux in linear scale on the y-axis. The time is given in Modified Julian Date (MJD)[8] which is the number of days elapsed since 17th of November 1858. The rainbow-color bar is linearly distributed[9] over the time span of each data set.

This representation aims to visualize stationarity within the observed period. Stationarity can be defined by the time invariance of particular statistical quantities. In a

---

[7] Utilized implementation: https://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.stats.ks_2samp.html, 17.03.2020; hereafter KS test
[8] See http://scienceworld.wolfram.com/astronomy/ModifiedJulianDate.html, 19.02.2020
[9] Taking 20 non-overlapping data points of the light curve with each step

A_g	μ_g	σ_g	χ^2_R	N_{bins}	skew	kurtosis
89.038	-0.067	0.259	0.0713	26	0.079	-0.336

μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_7	μ_8	μ_9	σ_μ
-0.211	-0.1	-0.011	-0.044	-0.089	-0.027	-0.068	-0.182	-0.146	0.004

Figure 2.3.: OJ 287; *Top:* Distribution of flux values F_i scaled by the mean $\overline{F}$ in logarithmic scale with Poisson errors (grey) and Gaussian fit (black). Fit parameters thereof and properties of the logarithmic distribution are shown in the upper table. *Bottom:* Light curve provides color coding according to time for the observation period of about 3600 days. A stationary distribution, is expected to scatter around the same mean for all colors, while non-stationary ones should show skewed accumulation of colors. For comparison, sationarity is indicated by the variance σ_μ of the nine mean values μ_i (computed from non-overlapping segments of 400 days) being relatively small.

PKS0208-512

A_g	μ_g	σ_g	χ^2_R	N_{bins}	skew	kurtosis
178.707	-0.112	0.243	0.091	14	-0.133	-0.476

μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_7	μ_8	μ_9	σ_μ
0.003	-0.191	-0.043	-0.173	-0.109	-0.13	-0.127	-0.011	-0.064	0.004

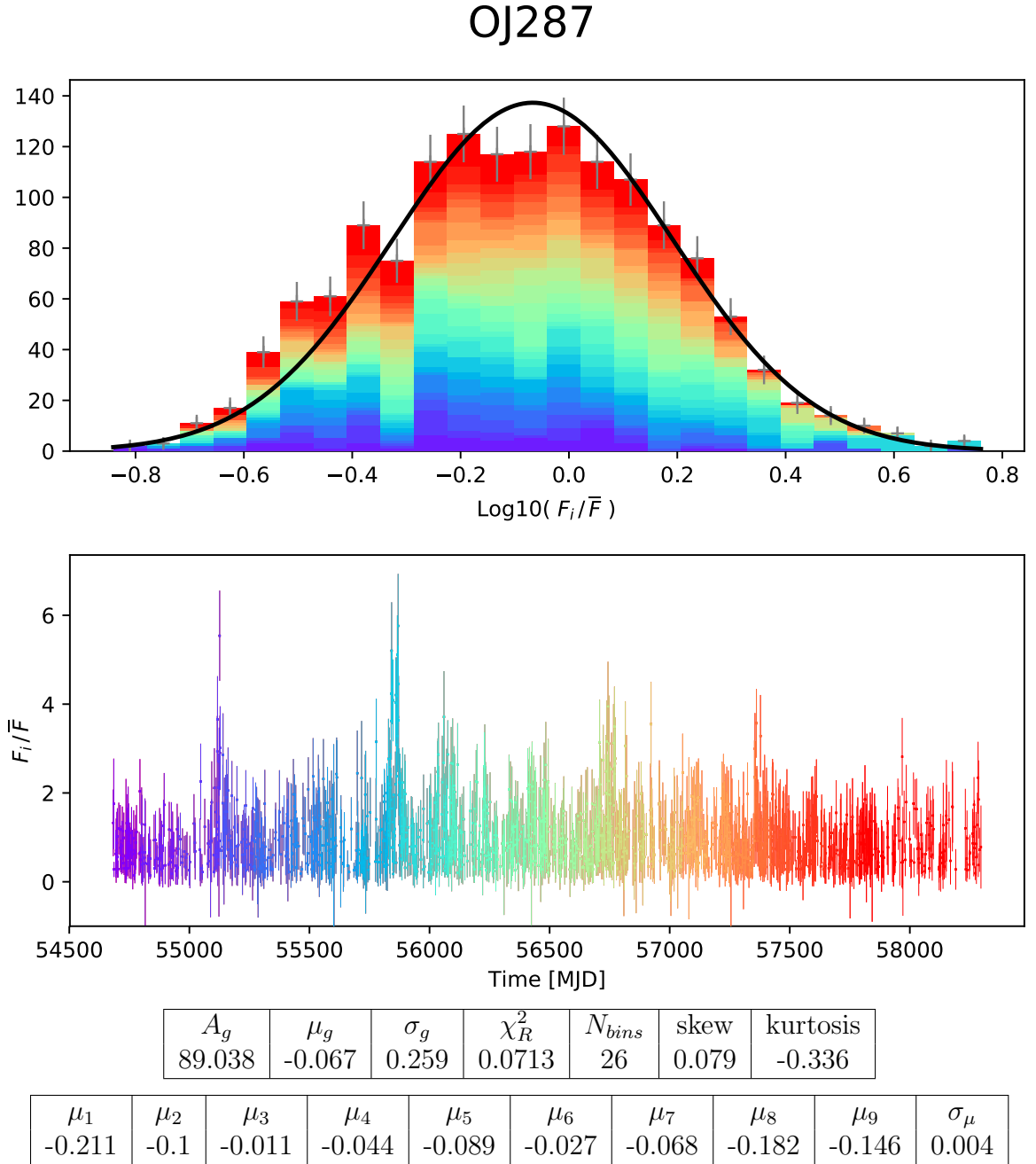**Figure 2.4.:** PKS 0208-512; *Top:* Distribution of flux values F_i scaled by the mean $\overline{F}$ in logarithmic scale with Poisson errors (grey) and Gaussian fit (black). Fit parameters thereof and properties of the logarithmic distribution are shown in the upper table. *Bottom:* Light curve provides color coding according to time for the observation period of about 3600 days. A stationary distribution is expected to scatter around the same mean for all colors, while non-stationary ones should show skewed accumulation of colors. Another indication of stationarity is that the variance σ_μ of the nine mean values μ_i (computed from non-overlapping segments of 400 days) is small.

PKS0235+164

A_g	μ_g	σ_g	χ^2_R	N_{bins}	skew	kurtosis
99.153	-0.085	0.318	0.092	25	0.039	-0.503

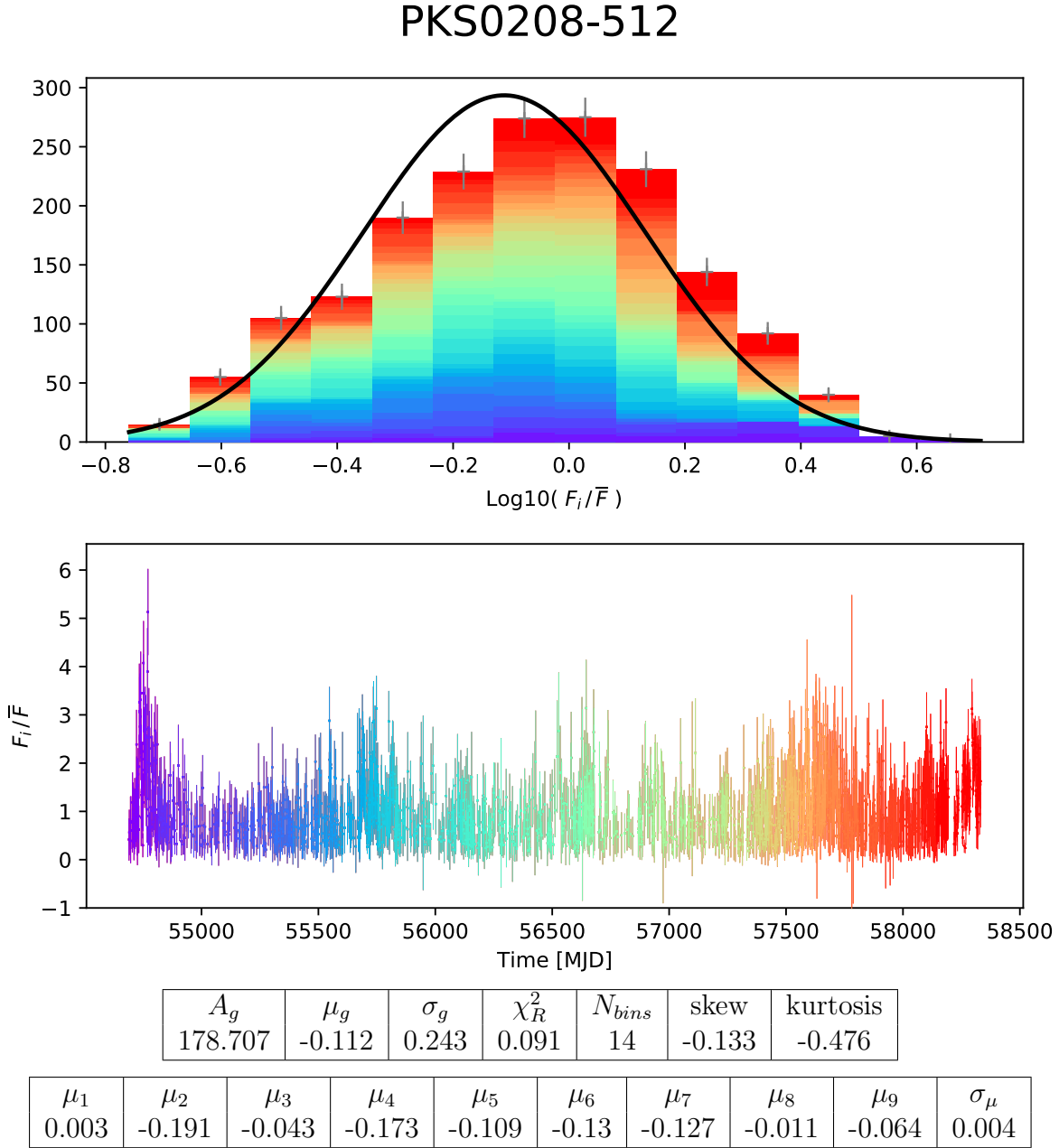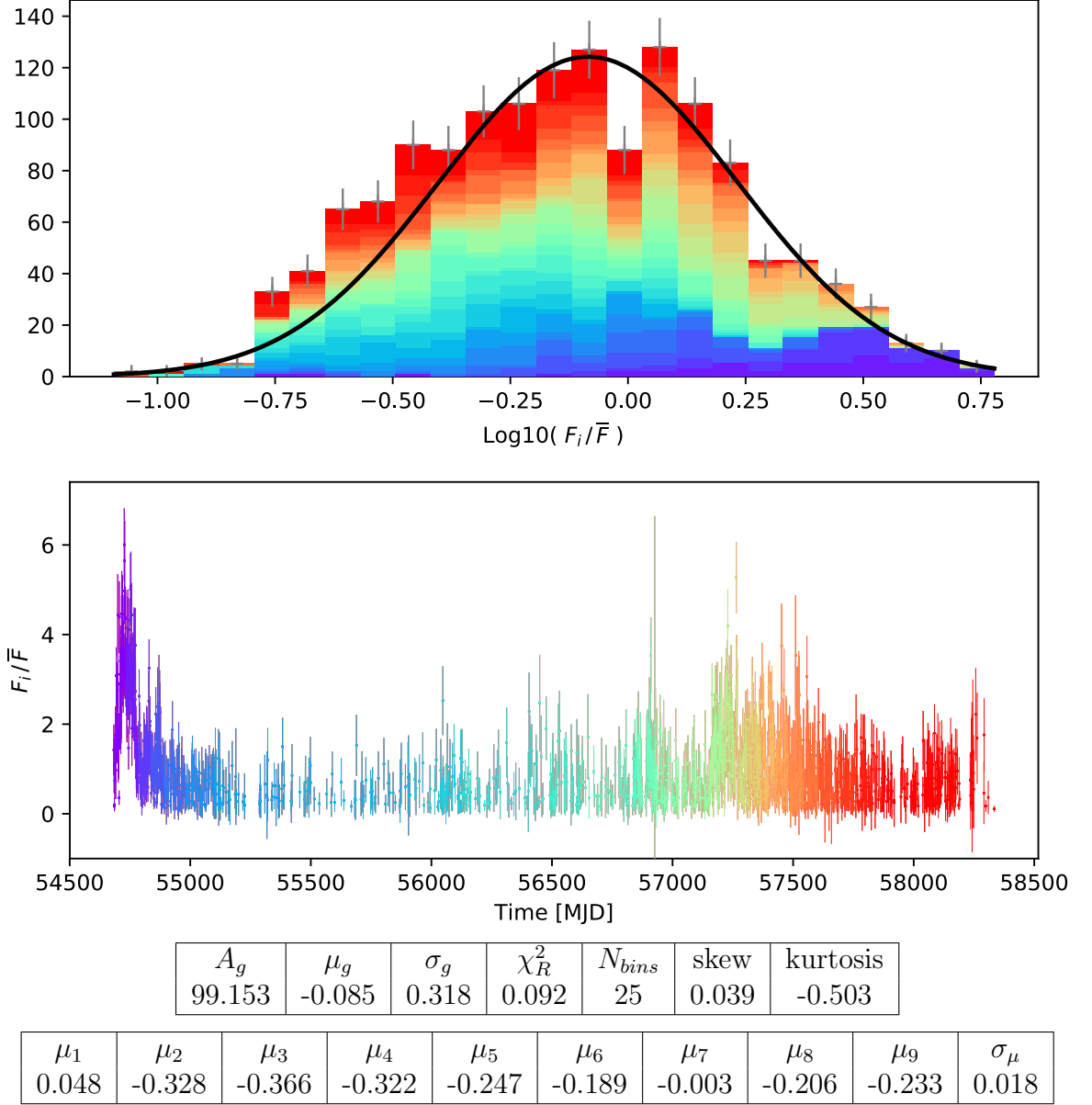| μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_7	μ_8	μ_9	σ_μ
0.048	-0.328	-0.366	-0.322	-0.247	-0.189	-0.003	-0.206	-0.233	0.018

Figure 2.5.: PKS 0235+164; *Top:* Distribution of flux values F_i scaled by the mean $\overline{F}$ in logarithmic scale with Poisson errors (grey) and Gaussian fit (black). Fit parameters thereof and properties of the logarithmic distribution are shown in the upper table. *Bottom:* Light curve provides color coding according to time for the observation period of about 3600 days. A stationary distribution is expected to scatter around the same mean for all colors, while non-stationary ones should show skewed accumulation of colors. Another indication of stationarity is that the variance σ_μ of the nine mean values μ_i (computed from non-overlapping segments of 400 days) is small.

PKS1510-089

A_g	μ_g	σ_g	χ^2_R	N_{bins}	skew	kurtosis
210.797	-0.169	0.379	0.190	42	0.087	0.170

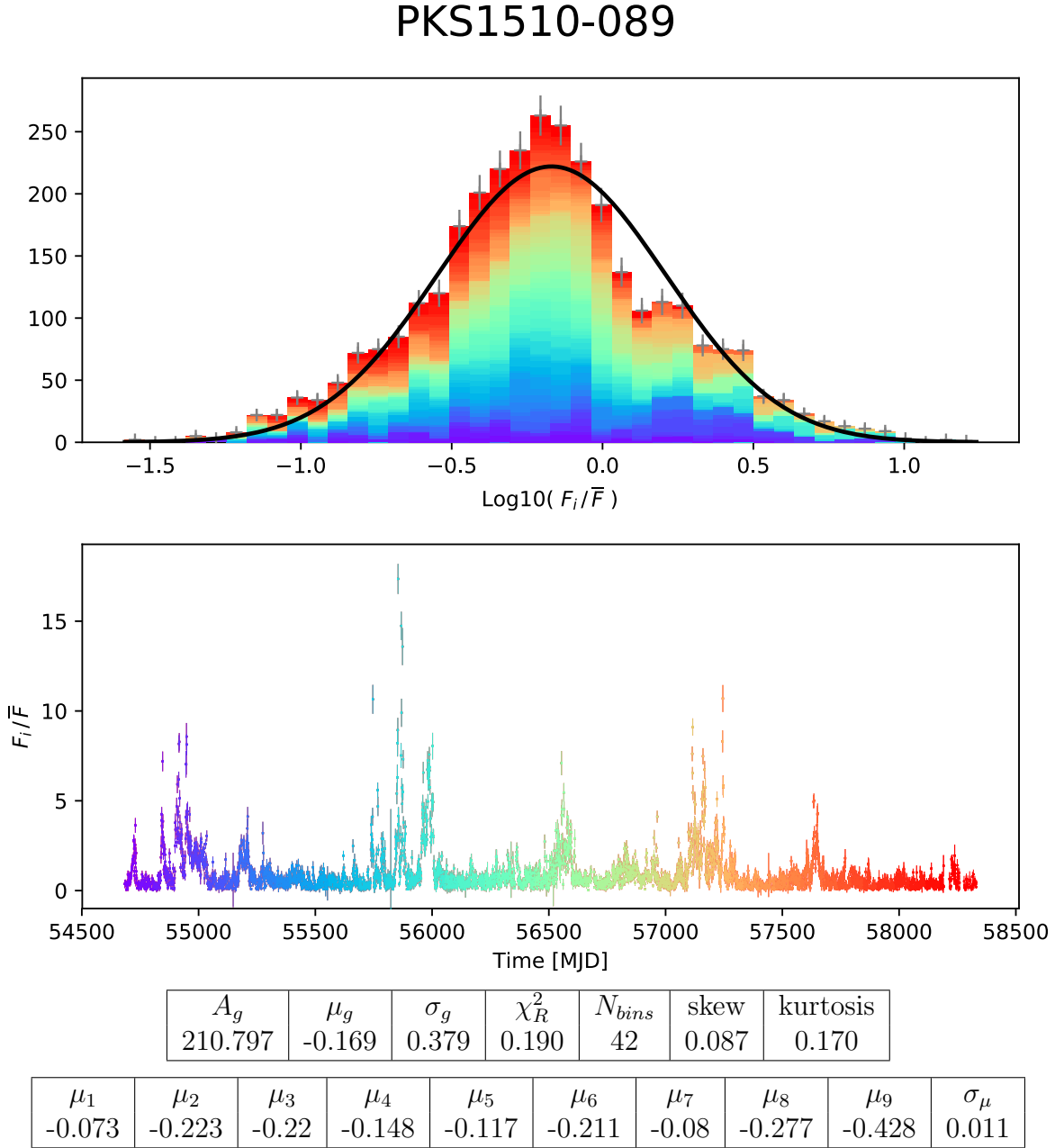| μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_7	μ_8	μ_9	σ_μ
-0.073	-0.223	-0.22	-0.148	-0.117	-0.211	-0.08	-0.277	-0.428	0.011

Figure 2.6.: PKS 1510-089; *Top:* Distribution of flux values F_i scaled by the mean $\overline{F}$ in logarithmic scale with Poisson errors (grey) and Gaussian fit (black). Fit parameters thereof and properties of the logarithmic distribution are shown in the upper table. *Bottom:* Light curve provides color coding according to time for the observation period of about 3600 days. A stationary distribution is expected to scatter around the same mean for all colors, while non-stationary ones should show skewed accumulation of colors. Another indication of stationarity is that the variance σ_μ of the nine mean values μ_i (computed from non-overlapping segments of 400 days) is small.

PKS2155-304

A_g	μ_g	σ_g	χ^2_R	N_{bins}	skew	kurtosis
123.346	-0.090	0.241	0.158	30	-0.068	-0.433

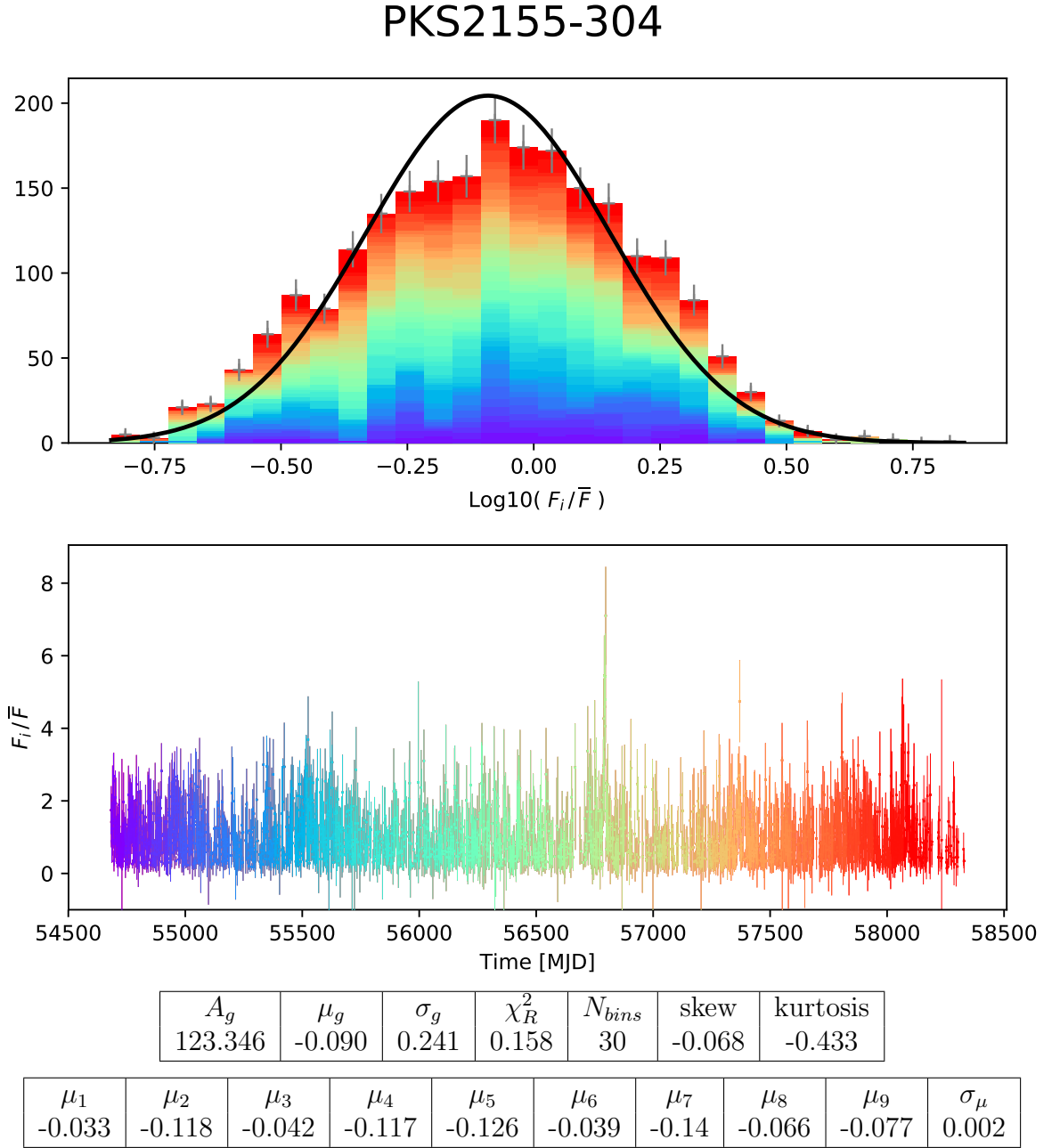| μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_7	μ_8	μ_9	σ_μ
-0.033	-0.118	-0.042	-0.117	-0.126	-0.039	-0.14	-0.066	-0.077	0.002

Figure 2.7.: PKS 2155-304; *Top:* Distribution of flux values F_i scaled by the mean $\overline{F}$ in logarithmic scale with Poisson errors (grey) and Gaussian fit (black). Fit parameters thereof and properties of the logarithmic distribution are shown in the upper table. *Bottom:* Light curve provides color coding according to time for the observation period of about 3600 days. A stationary distribution is expected to scatter around the same mean for all colors, while non-stationary ones should show skewed accumulation of colors. Another indication of stationarity is that the variance σ_μ of the nine mean values μ_i (computed from non-overlapping segments of 400 days) is small.

A_g	μ_g	σ_g	χ^2_R	N_{bins}	skew	kurtosis
115.278	-0.071	0.321	0.263	42	0.324	0.447

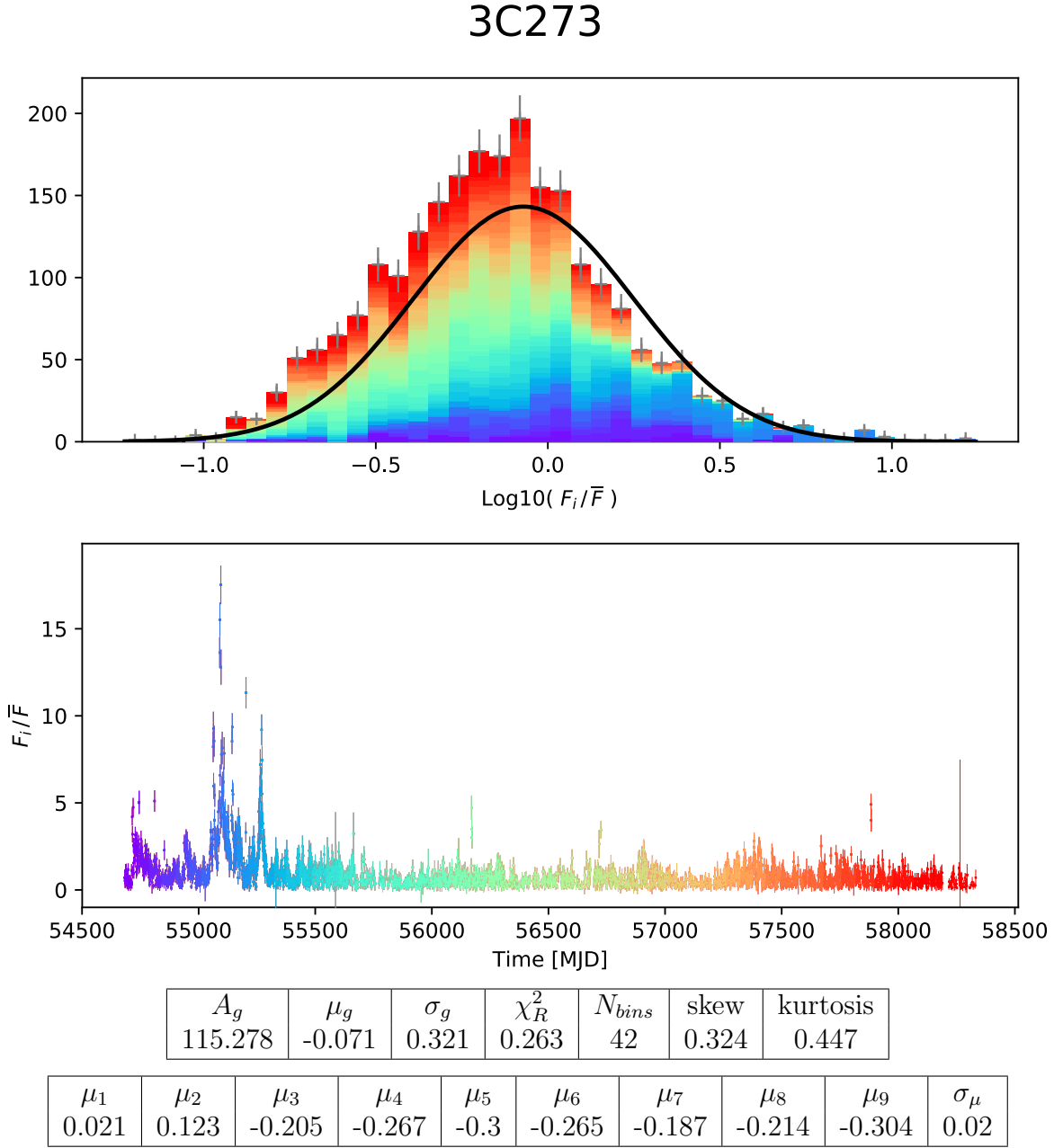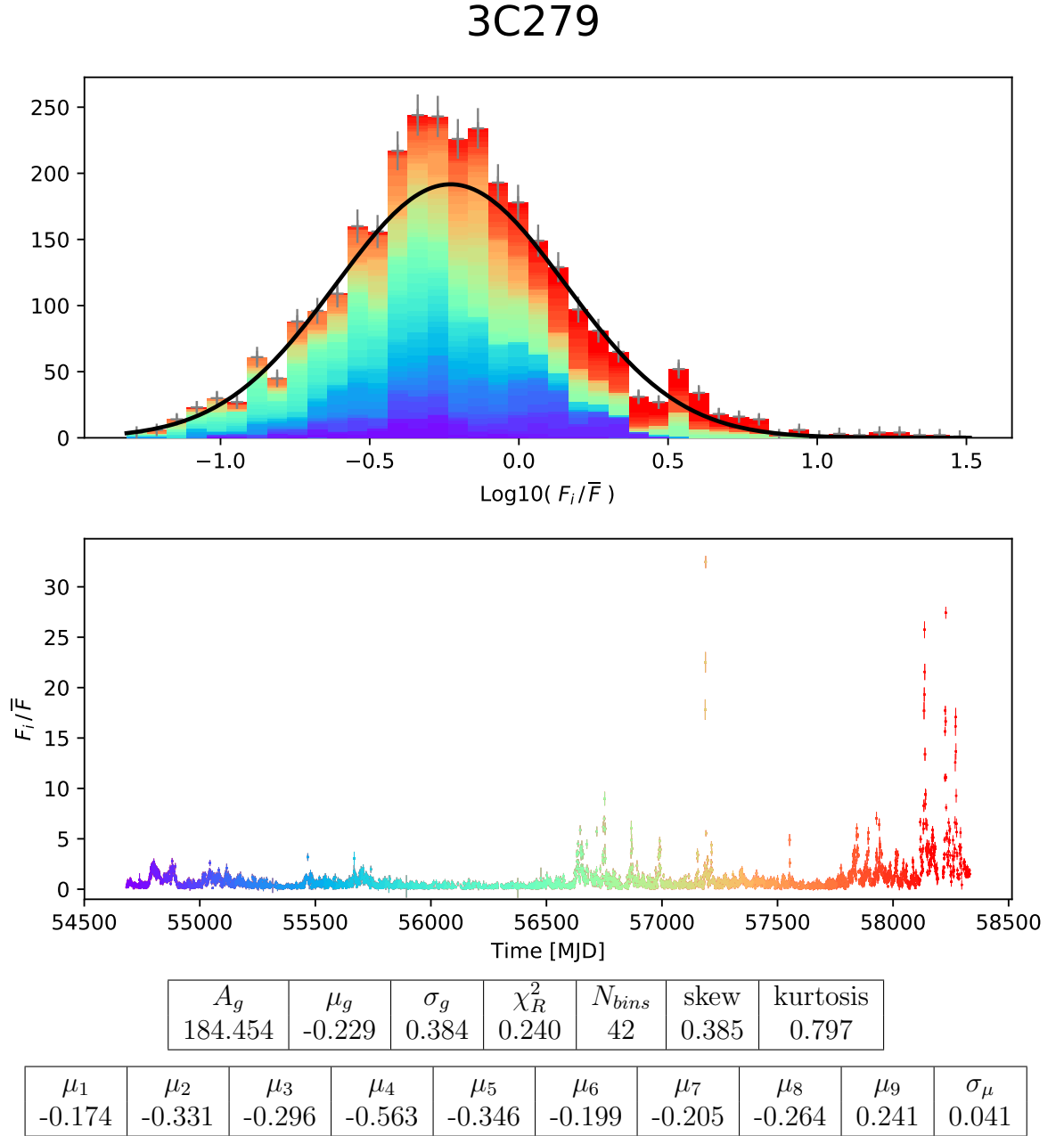| μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_7	μ_8	μ_9	σ_μ
0.021	0.123	-0.205	-0.267	-0.3	-0.265	-0.187	-0.214	-0.304	0.02

Figure 2.8.: 3C 273; *Top:* Distribution of flux values F_i scaled by the mean $\overline{F}$ in logarithmic scale with Poisson errors (grey) and Gaussian fit (black). Fit parameters thereof and properties of the logarithmic distribution are shown in the upper table. *Bottom:* Light curve provides color coding according to time for the observation period of about 3600 days. A stationary distribution is expected to scatter around the same mean for all colors, while non-stationary ones should show skewed accumulation of colors. Another indication of stationarity is that the variance σ_μ of the nine mean values μ_i (computed from non-overlapping segments of 400 days) is small.

3C279

A_g	μ_g	σ_g	χ^2_R	N_{bins}	skew	kurtosis
184.454	-0.229	0.384	0.240	42	0.385	0.797

μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_7	μ_8	μ_9	σ_μ
-0.174	-0.331	-0.296	-0.563	-0.346	-0.199	-0.205	-0.264	0.241	0.041

Figure 2.9.: 3C 279; *Top:* Distribution of flux values F_i scaled by the mean $\overline{F}$ in logarithmic scale with Poisson errors (grey) and Gaussian fit (black). Fit parameters thereof and properties of the logarithmic distribution are shown in the upper table. *Bottom:* Light curve provides color coding according to time for the observation period of about 3600 days. A stationary distribution is expected to scatter around the same mean for all colors, while non-stationary ones should show skewed accumulation of colors. Another indication of stationarity is that the variance σ_μ of the nine mean values μ_i (computed from non-overlapping segments of 400 days) is small.
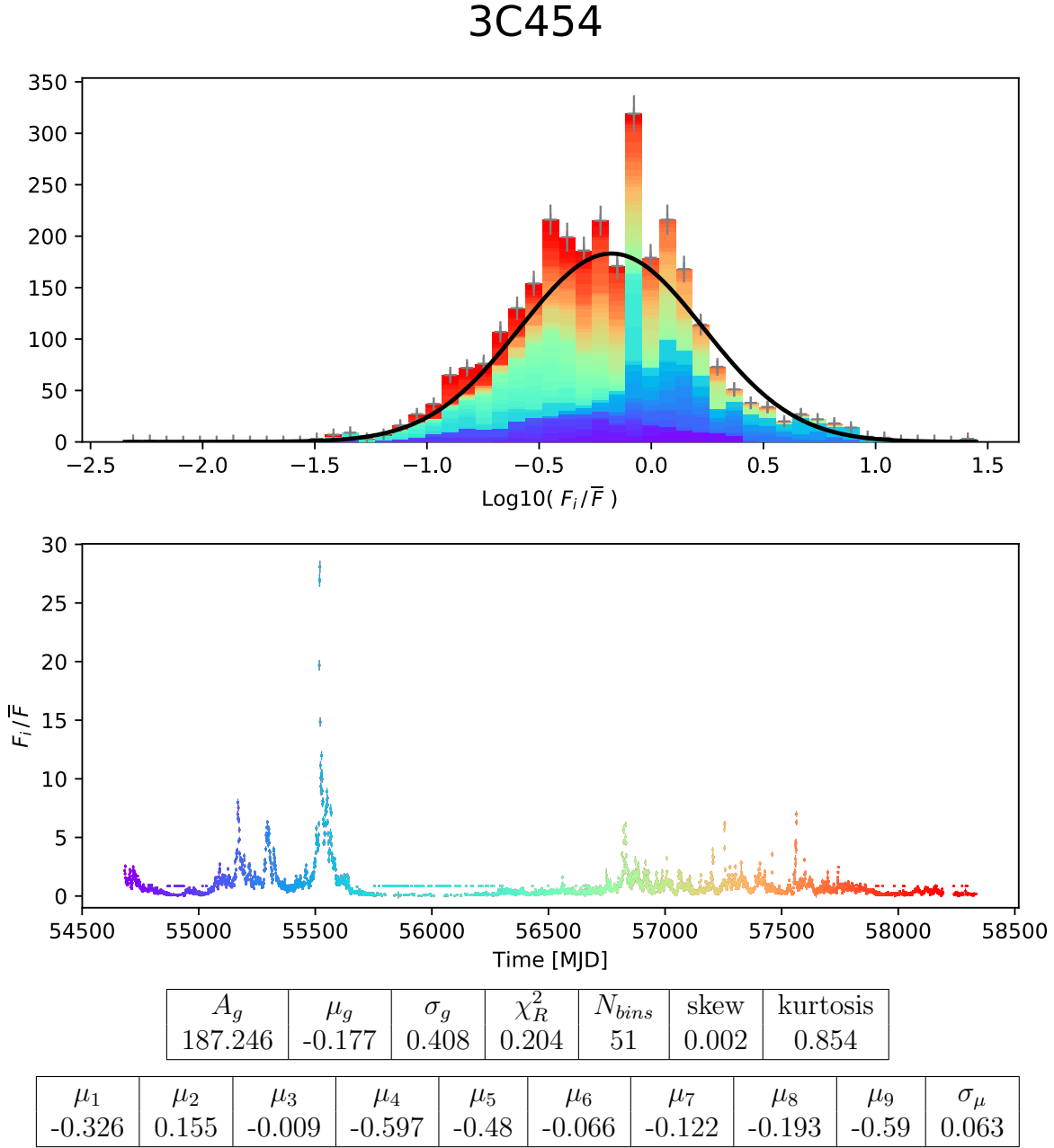
A_g	μ_g	σ_g	χ^2_R	N_{bins}	skew	kurtosis
187.246	-0.177	0.408	0.204	51	0.002	0.854

μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_7	μ_8	μ_9	σ_μ
-0.326	0.155	-0.009	-0.597	-0.48	-0.066	-0.122	-0.193	-0.59	0.063

Figure 2.10.: 3C 454.3; *Top:* Distribution of flux values F_i scaled by the mean $\overline{F}$ in logarithmic scale with Poisson errors (grey) and Gaussian fit (black). Fit parameters thereof and properties of the logarithmic distribution are shown in the upper table. *Bottom:* Light curve provides color coding according to time for the observation period of about 3600 days. A stationary distribution is expected to scatter around the same mean for all colors, while non-stationary ones should show skewed accumulation of colors. Another indication of stationarity is that the variance σ_μ of the nine mean values μ_i (computed from non-overlapping segments of 400 days) is small.

wide sense, stationarity is defined by the invariance of the mean, a finite variance, and the fact that the co-variance only depends on the distance between two data points. Theoretically, the time invariance could only be checked if one had an infinite amount of data points. The finite observations acquired in practice therefore have to be considered with care: in a specific stretch of data, a stationary distribution could be mistaken to be non-stationary and vice versa (Scargle, 2020). Bearing this in mind, it shall be investigated whether the scaled, daily log-fluxes are stationary within the observation period of about 3600 days.

For the illustrative purpose of this work, indications of stationarity are studied through the time invariance of the mean by dividing each light curve in segments of 400 days. This results in nine segment-mean values μ_i that are shown in the bottom table of Fig. 2.3 - 2.10 along with their variance σ_μ.
The variance of the segment-mean values is used as estimate of stationarity where a small value with respect to the segment-mean values indicates time invariance (with respect to the segment size). This is visualized in the histograms through color coding according to time of the light curve. A hint for stationarity is imprinted by the fact that all colors scatter about the same, time invariant mean i.e. are distributed equally over the span of the histogram. This is fulfilled for the histograms of OJ 287, PKS 0208-512, and PKS 2155-304. In agreement with that, the variance σ_μ of the mean yields small (< 0.004) values. High variance values (> 0.40) suggest that the distribution is not stationary within the observation period with respect to the segment size. This is the case for 3C 279 and 3C 454.3. Accordingly, the colors in the histogram do not scatter about the same value but show distinct clustering. For 3C 279, mainly the last, red part appears shifted with respect to the previous flux mean. 3C 454.3 shows the highest values of σ_μ and consequently distinct accumulation of colors in the histogram. It can be seen that the blueish phase around MJD 55500 is at a higher flux level than the turquoise-green, quiescent phase between 56000 and 56500. PKS 0235+164 yields an intermediate variance of 0.018. In the histogram it can be verified that this is because the distribution seems to be stationary apart from the first, purple segment around MJD 54750 which appears to scattered around a higher flux level. This instance shows how much the selection of the observation period can affect the judgment of stationarity. The remaining sources PKS 1510-089, and 3C 273 yield intermediate variance values and can neither by this value nor by the color coded histogram be classified to be stationary or

not. As already mentioned, this only holds for the considered time segments of 400 days within the observation period of about 3600 days.

In the next step, the logarithmic flux distributions plotted in Fig. 2.3-2.10 are fitted with a Gaussian function of the from

$$y = \frac{A_g}{\sqrt{2\pi} \cdot \sigma_g} \cdot \exp\left(\frac{-(x - \mu_g)^2}{2 \cdot \sigma_g^2}\right) \tag{2.1.2}$$

which corresponds to a log-normal fit of the linear flux data.

When presenting fits, it is crucial to give an estimate for the goodness of the fit. To do so, one first needs to determine the uncertainties of the data points to be fitted. In this case, this corresponds to the uncertainty σ_b of counts in each bin of the histogram. According to Poisson statistics, this equals $\sigma_b = \sqrt{n_b}$ for n_b counts in each bin which is shown in the plots with grey error bars. When fitting[10] the Gaussian above to the histogrammed data, this uncertainty is considered by setting the weights to $1/\sigma_b^2 = 1/n_b$. The best fit for the free parameters amplitude A_g, mean value μ_g, and variance σ_g of Eq. 2.1.2 are shown in the table of Fig. 2.3-2.10. As measure for the goodness of the fit, the reduced chi square χ_R^2 and the corresponding number of fitted data points (i.e. number of bins N_{bins} determined with the Knuth algorithm) is shown as well. All of the fit parameters and the third[11] and fourth[12] statistical moment are discussed in Sec. 3.1 and summarized there in Tab. 3.1.

For each of PKS 0208-512, PKS 1510-089, and 3C 273, the fit seems to be shifted. This effect can be explained by the uncertainty of each histogram bin, which has been estimated with Poisson statistics. The dependence of uncertainty σ_b on the counts n_b causes the histogram bins with only few counts to have a relatively small error while the ones with many counts have a large error. Consequently, the small histogram bins will be weighted stronger than the larger ones.

The histogram of 3C 454.3 shows one particularly high bin that seems to break the distribution. A closer look at the light curve shows that this is probably due to the part in the light curve where the flux seems to vary between two fixed values. Inspection of

---

[10] Utilized implementation: https://lmfit.github.io/lmfit-py/model.html, 20.02.2020

[11] Utilized implementation: https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.skew.html, 04.03.2020

[12] Utilized implementation: https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.kurtosis.html, 04.03.2020

the data reveals that these flux values have unexpectedly high TS values (TS > 3000). Based on this, these data points are assumed to be artifacts of the analysis.

For an ideal fit, the reduced chi square χ^2_R should be close to one. The fact that $\chi^2_R \ll 1$ for all fits suggests that the error of the histogram bins have been considered too large. Instead of using Poisson statistic, the option of propagating the error of the flux values on the error of the counts in the histogram, similar to Meyer et al. (2019), shall be explored in future work.

2.2. Definition of flares in light curves

There are two approaches to characterize and model the variability of observed flux variations. They differ in that one makes physical assumptions, while the other does not. They are thus referred to as physically motivated (physical) and purely data driven (non-physical) models, respectively. Physical models assume a certain shape and behavior of the flux given a theoretical model and its implications on e.g. acceleration, cooling, dynamical, and instability timescales. In this work, the goal is to put constraints on the possible physical processes. Therefore, time domain modeling of the flux should be non-physical (i.e. not biased towards a physical model).

Besides auto-correlation and Fourier analysis, Bayesian Block analysis (Scargle et al., 2013) is frequently used for this purpose and chosen here. It allows one to "identify and characterize statistically significant variations while suppressing the inevitable corrupting observational errors." The procedure is illustrated by Fig. 2.11, which shows the monthly light curve of PKS 1510-089. The top panel shows the "raw" light curve as obtained by *Fermi*-analysis. The middle panel shows the corresponding Bayesian blocks as obtained by the implementation outlaid in A.1, which is used for all of the results throughout this thesis.

For the execution of this algorithm, the fitness is set to *measures*, assuming that the flux errors are Gaussian[13]. It is possible to adjust the p-value or the gamma-value of the algorithm, which affects the fineness of the blocks. While the robustness of the methodology was tested with various settings of these parameters, the results presented here are produced with default settings only. The flux value of a block is set to the mean of
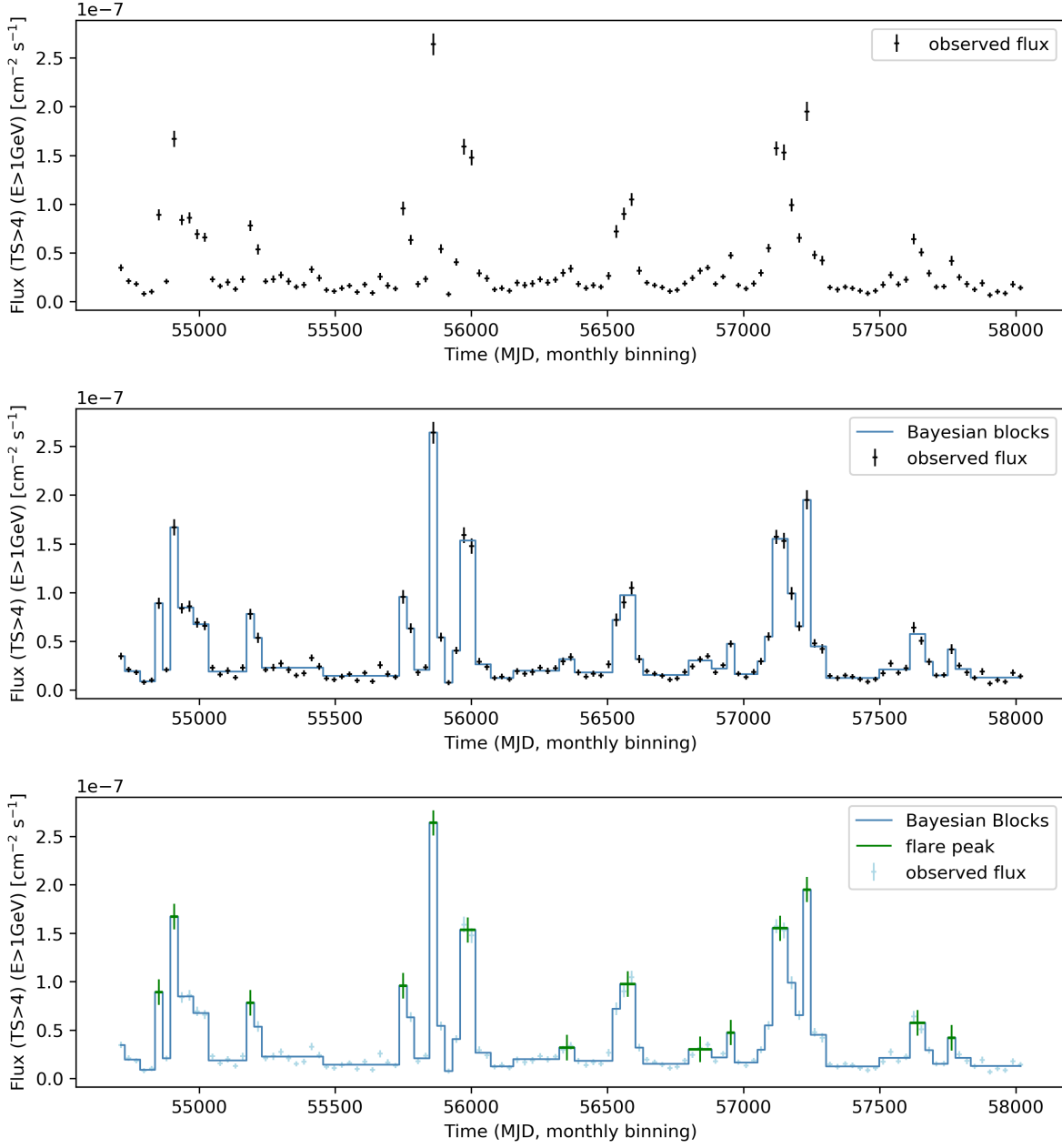
---

[13] Strictly, these flux errors are Possonian.

Figure 2.11.: Monthly light curve of PKS 1510-089 (top) analyzed with Bayesian blocks shown in blue (middle). The center of the local block-maxima are defined to be individual peaks of flaring events as indicated with a green vertical line (bottom). This is the starting point for both flare finding methods.

the data points within the block. Though this is by far the most common application, the algorithm itself does not account for the shape or the flux value of the blocks. In future work, this could be explored by setting the block to other values (e.g. median, minimum, or maximum) or fitting the data within differently.

Another factor that influences the outcome of the Bayesian block algorithm is the data itself, of course. As explained in the previous section, flux values violating the condition TS > 4 are neglected here. It has to be noted that taking these values or their upper limits in account can result in different blocks. Furthermore, the light curves analyzed here are binned monthly or daily which requires an arbitrary time to be the start of a month or a day. It has been shown that shifting the integration time of daily binning by 12 or 6 hours results in different bins and consequently affects the block representation (A. Gokus, private communication). Consideration of these effects, however, exceeds the scope of this work.

Keeping these limitations in mind, the blocks determined with default settings provide the basis for the definition of a flare. This is achieved with two methods (baseline and half-clap) that are introduced in the following. For both of the methods, the first step is to determine the local maxima of the blocks in the light curve. Every block that is higher than its two neighbors is considered to be the peak of a flare. Referring to the definition of the Bayesian block algorithm, there is no significant variation within a block. Hence, the exact time of these flare peaks is set to the middle of the block, as illustrated in Fig. 2.11 in the bottom panel with vertical lines over the peak block in green.

The next step in both methods is to define which blocks belong to the rise and decay of these peaks. This is done by proceeding downward from each peak block, as long as the adjacent block is lower. One can imagine this like pouring water onto a peak. All the blocks that the water is is running over, belong to this peak and therefore form a flare. This is equal to the a one-dimensional version of the watershed concept of topological data analysis. The natural follow up question is: what about the resulting lakes or valley blocks? I.e. where does one flare end and the other start? This is where the two definitions of flares of this work provide different approaches.

2.2.1. Baseline Method

The baseline method[14] was first introduced in Meyer et al. (2019). The authors use the criteria of a constant flux level that defines the onset and offset of a flare if the flux exceeds or goes under that level, respectively. This means only peaks that are above this baseline are considered as flare and adjacent blocks belong to a peak, as long as they are above the baseline. Meyer et al. (2019) use the mean flux of each light curve as baseline which is adapted here. The baseline method is illustrated in Fig. 2.12 in the top panel. In addition to the identified peaks represented in green, the baseline is shown in pink. The block groups that consequently define a flare are shaded in orange. The authors point out that this approach stems from the so-called HOP algorithm (Eisenstein & Hut, 1998) as in "hopping" from one point (block) to the other. This is why the resulting flares are referred to as HOP-groups. For the scope of this work, multiple peaks are treated as one flare, as shown in the third flare from the right in the top panel of Fig. 2.12. The highest local maximum within the flare represents the peak. The first (last) flare in the light curve starts (ends) when the flux rises above (goes under) the baseline the first (last) time. This means flares that lack an onset or offset or cases where the flux starts off or ends above the baseline are conservatively dismissed for this analysis.

The clear disadvantage of this method is that the result is highly dependent on the choice of baseline. Moving it further up or down will exclude or include different groups of blocks which in turn highly affects the shape of the flare. Moreover, multiple maxima within one flare defined in this way could represent what are in fact multiple, distinct flares as one single flare. What appears to be a quiescent level of the source by eye, i.e. the comparably long low-flux blocks, can thus not serve as baseline to define the flares with this method. The higher the baseline is set the more peaks can possibly lay under the baseline and thus not be considered at all. A clear advantage of this method is that despite the baseline being a bias, it is the only necessary assumption to systematically define flares in every light curve.
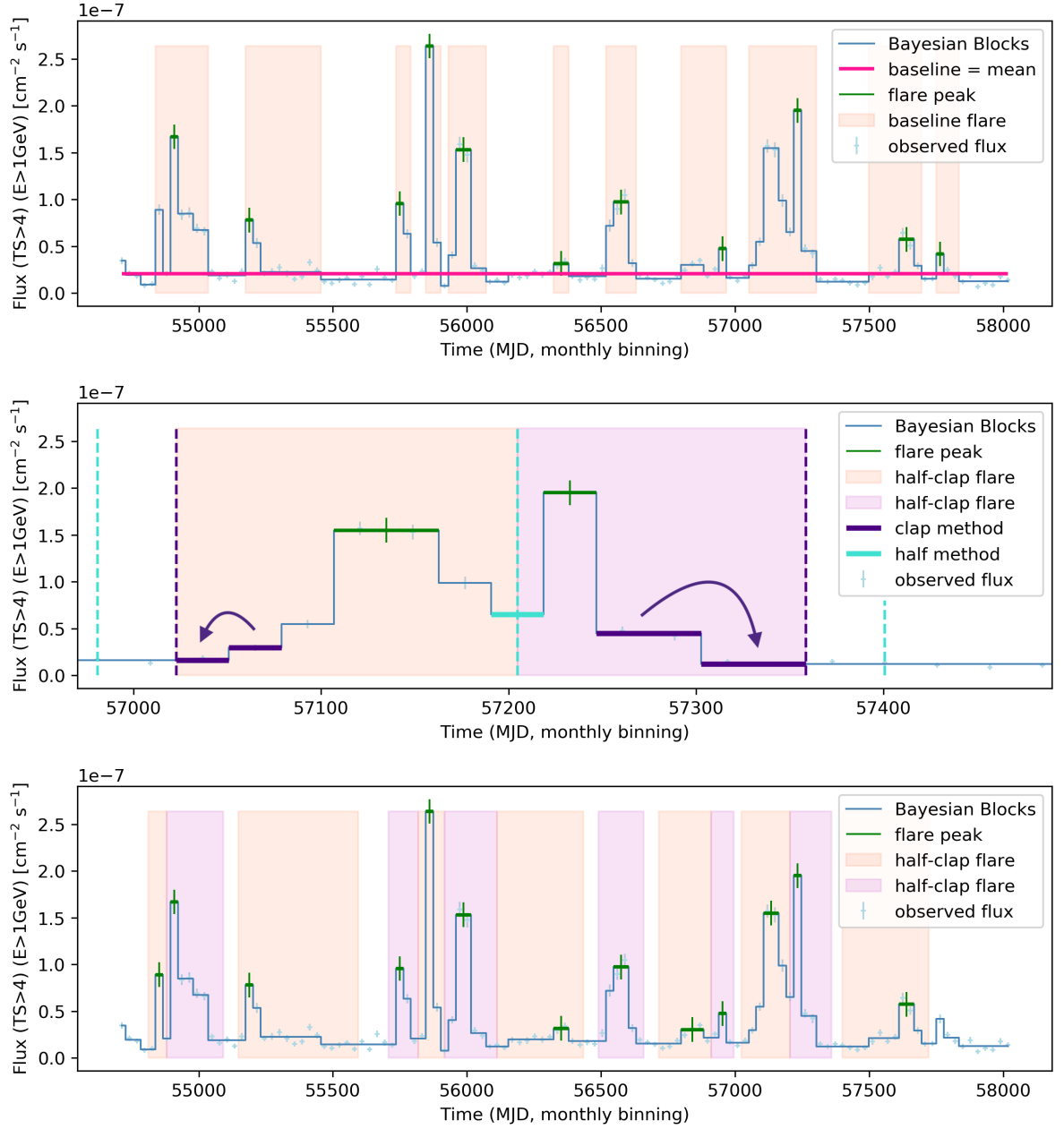
---

[14] Implementation: see Appendix A.2

Figure 2.12.: Monthly light curve of PKS 1510-089 and identified flares based on the baseline method (top) and the half-clap method (bottom). For the former, the start and end of a flare is defined by blocks going over or under the baseline. In case of multiple peaks (e.g. the third one from the right) the highest peak is chosen. The half-clap method defines the start and end of a flare by taking the minimum of half the valley block (turquoise) and clapping the adjacent block onto the valley block (dark purple), as illustrated in a zoomed in section of the light curve (middle).

2.2.2. Half-Clap Method

The half-clap method[15] has been established in the framework of this work to avoid the bias of choosing a baseline. The first step, is to divide the mentioned valley blocks analogous to the peak blocks at their center (half method). This way, all of the light curve could be divided into flares based on the local maxima and minima of the blocks. There are, however, many cases where the valley block is very long. This corresponds to a quiescent flux state without flaring from the source rather than to the onset or offset of a rapid flux variation. In these cases, dividing the valley block at its center would highly exaggerate the extent of the flares. Thus, the slope of the variation is taken into account. This is done by clapping the length of the block adjacent to the valley block onto that valley block (clap method). The resulting fraction of the valley block is then interpreted to be part of the flare whereas the rest of the valley block represents a quiescent level. The methods are combined by taking the minimum of each method for the start and end of a flare. This basically means that the start (end) of a flare lays at the center of a valley block unless the rising (falling) behavior suggests that it should be closer to the peak.

This method is illustrated in the middle panel of Fig. 2.12. It shows the same blocks that represent the third flare from the right of the top panel in the same figure[16]. The two flares shaded in orange and purple in the middle panel were defined based on the half-clap method. First, the center of each valley block is considered to be the end of one and the start of another variation. This is illustrated with a turquoise, dashed line (half method). Second, the rising and falling behavior is analyzed by clapping the length of the adjacent block onto the valley block. This is shown with dark purple for the beginning of the left and the end of the right flare in the middle panel of Fig. 2.12.

For the onset of the left, orange flare, the rising behavior determined with the clapping method yields a shorter rise time (starting slightly over MJD 57000, shown in dark purple) than the half method does (starting slightly under MJD 57000, shown in turquoise). Therefore the minimum (clap method in this case) defines the onset of the left flare. For the end of that flare, by contrast, the half method (ending just over MJD 57200, shown

---

[15] Implementation: see Appendix A.3

[16] While the baseline method shown in the top panel, identifies one flare, the half-clap method in the middle panel identifies two flares for the same block structure. The two independent methods to define flares naturally yield different flares for the same light curve.

in turquoise) represents the minimum. The same holds for the onset of the right flare shaded in purple. Consequently, the left flare ends where the right flare starts. As can be seen by this example, the defined flares do not overlap and are uniquely identified within the light curve. Relaxing this condition is part of future work. The end of the right flare is again given by the clap method (ending at MJD 57350, shown in dark purple) as it is shorter than the half method (ending at MJD 57400, shwon in turquoise).

Applying this procedure to the full light curve results in the flares shaded in orange and purple in the bottom panel of Fig. 2.12. Also for this method, flares that are not completely mapped (e.g. missing a start or an end) are conservatively neglected. The last group of blocks in the bottom panel of Fig. 2.12, for instance, is not considered because it is not certain how much further the last block will go and whether the next block would be higher or lower than that.

Even though this method does not require a baseline, assumptions have to be taken into account to define flares. One strong assumption is that the slope of the flare can simply be extended by clapping the adjacent block. The second assumption is that the length of half the valley block belongs to the flare instead, in case this length is less than the clapping would be. The latter represents handling the special case of overlapping flares based on the former. While these strong assumptions are a downside compared to only having one assumption in the baseline method, this approach allows to consider every peak that is found in the light curve. This includes peaks with low flux values and close peaks at a high level, both of which are not considered with the baseline method.

2.2.3. Parametrization of the flares

The key question that shall be answered by systematically applying these methods to many light curves is whether the flares show a tendency for asymmetry based on the rise time t_r and the decay time t_d. These are simply given by the time between the onset of a flare and its peak and the time from the peak to the offset, respectively. One has to keep in mind that both of these methods are built up on non-physical time domain modeling. They do yield exact times but these are not necessarily equal to the physical rise and decay behavior. Nevertheless, the methods provide an objective parametrization that is
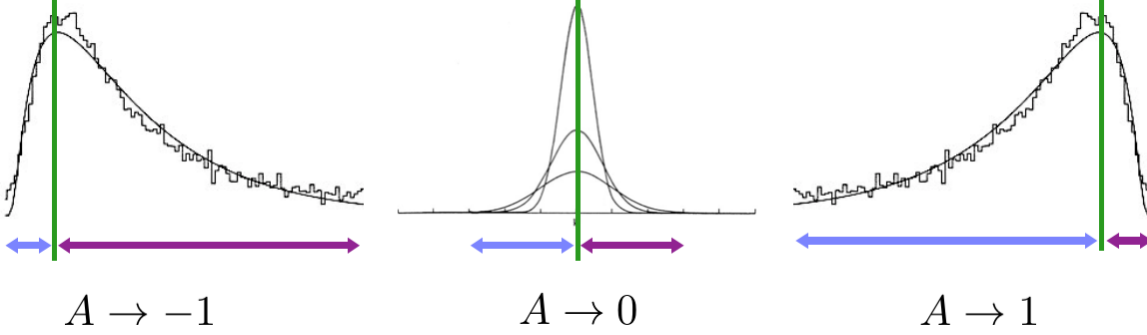
Figure 2.13.: Illustration of the asymmetry measure A (Eq. 2.2.1) for three exemplary flare shapes. The flare peak is denoted with a green vertical line and the rise and decay times are indicated with lavender and fuchsia horizontal arrows respectively.

studied in form of the asymmetry measure

$$A = \frac{t_r - t_d}{t_r + t_d} \tag{2.2.1}$$

of each flare. The three exemplary flare shapes in Fig. 2.13 illustrate that the asymmetry measure ranges from -1 to $+1$. In the following, all flares from all sources (classes) are considered at once as one large sample. In figures, the baseline method is always shown on the left side and the half-clap method on the right.

2.3. Monthly Flares

Both of the methods are applied to the monthly sample. As discussed in Sec. 2.1.1, this consists of 335 sources. For 99 of these sources, the Bayesian block algorithm yields only one block, i.e. the flux over the observed period is constant. Requiring that the light curve shows at least one full flare, the baseline method detects 537 flares in 196 sources and the half-clap method detects 401 flares in 119 sources. The discrepancy is due to the conservative conditions of the definition of a full flare explained in the previous section. The resulting distribution of asymmetry measures (see Eq. 2.2.1) $\tilde{A}_{m,bl}$ and $\tilde{A}_{m,hc}$ for all flares in all sources of the monthly sample is shown in the top left and right of Fig. 2.14 for the baseline and the half-clap method, respectively.

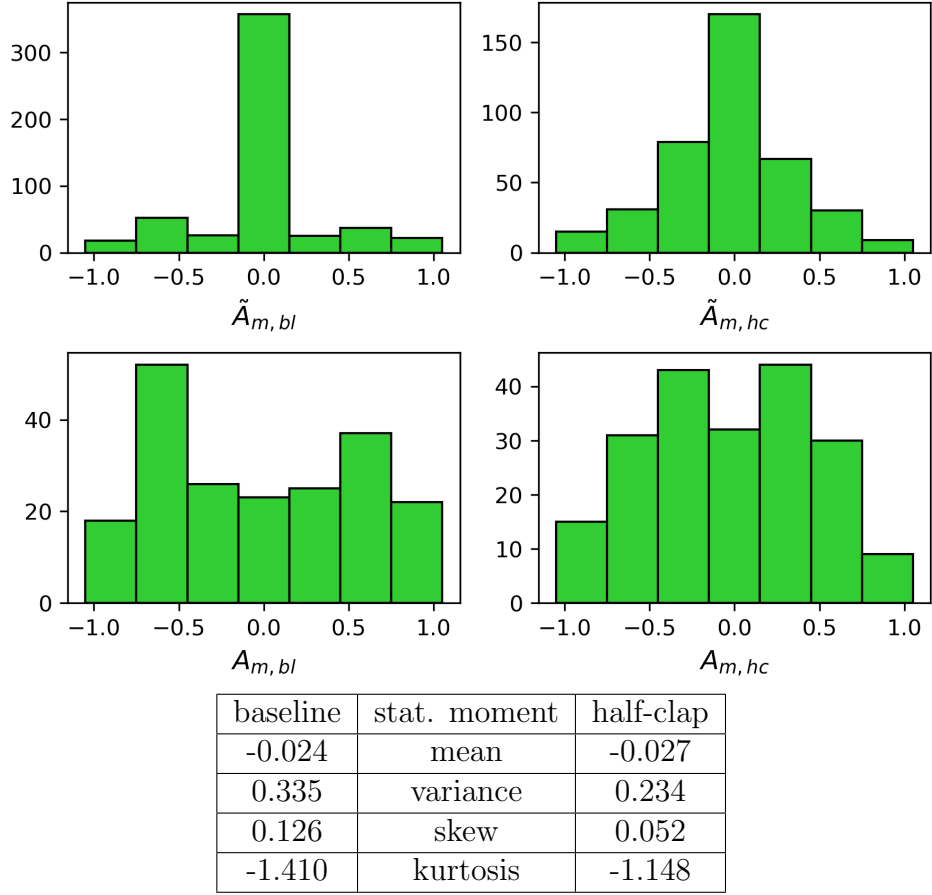Both methods show an excess for $A = 0$, suggesting that a large portion of all flares

baseline	stat. moment	half-clap
-0.024	mean	-0.027
0.335	variance	0.234
0.126	skew	0.052
-1.410	kurtosis	-1.148

Figure 2.14.: Distribution of asymmetry measures for all detected flares of all sources in the monthly sample derived with the baseline method (top left) and half-clap method (top right). Both methods yield each kind of asymmetry to a comparable degree when excluding single-block flares (bottom), as shown by the statistical moments of these distributions listed in the table.

are symmetrical. An investigating of the light curves, however, reveals that many of these symmetric flares consist of one block only. The sixth flare determined with the baseline method in the top panel of Fig. 2.12 is such an instance. A single block is not sufficient to represent the structure of a flux variation, as discussed in Sec. 3.2.4. Hence, single block flares are excluded for this analysis. For the half-clap method, this also includes flares such as the fifth one from the right in the bottom panel of Fig. 2.12. The asymmetry measure does not necessarily have to be zero, but if there is only one block (the peak) between the starting and end block, the flare is conservatively dismissed.

Only considering the asymmetry measure $A_{m,bl}$ and $A_{m,hc}$ of flares with more than one block yields 203 baseline flares and 204 half-clap flares as shown in the distributions on the left and right of the bottom panel in Fig. 2.14, respectively. Comparison of this to the result in the top panel shows that the excess of symmetric flares is a relic of non-resolved single-block flares rather than a real symmetry of the flux variations. Hereafter, all single-block flares will be excluded to avoid this bias.

Given this, it can be seen that the asymmetry measures for both methods range from -1 to 1. This means that for the monthly sample, both methods yield flares with both kinds of asymmetry as well as symmetry. For the baseline method, it appears that the second from left and second from right bin occur slightly more often than the rest. This could be due to an excess of flares that consist of only two blocks based on one data bin each. In this case, either the rise time is exactly twice as high as the decay time or vice versa (see Sec. 3.2).

The statistical moments (see Eq. 2.1.1) of both distributions without single-block flares are presented in Fig. 2.14 as well. The mean of both distributions is lower than zero, suggesting that flares $t_r < t_d$ occur slightly more often than the opposite. However, the deviation of each mean from zero is not significant with respect to the large variation of either method.

The skewness or asymmetry of both distributions is close to zero. This shows that asymmetry measures larger and smaller than zero occur to a comparable degree. It can be noted that the skewness of both methods is larger than zero, hinting a positive skew. Just as the tendency of the mean, this suggests that that flares with $t_r < t_d$ are slightly more numerous than the opposite. Neither of these results is significant, though.

The kurtosis of both distributions is less than zero and therefore boxier than a Gaussian. As can already be seen in the histograms, the clear deviation of the kurtosis from
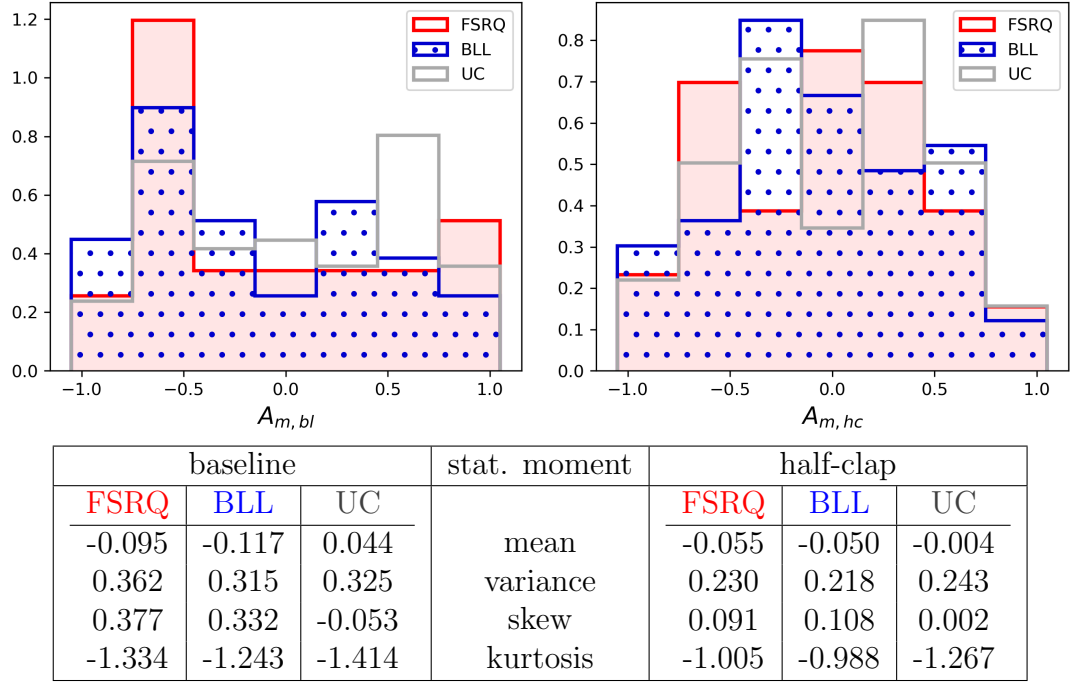
baseline			stat. moment	half-clap		
FSRQ	BLL	UC		FSRQ	BLL	UC
-0.095	-0.117	0.044	mean	-0.055	-0.050	-0.004
0.362	0.315	0.325	variance	0.230	0.218	0.243
0.377	0.332	-0.053	skew	0.091	0.108	0.002
-1.334	-1.243	-1.414	kurtosis	-1.005	-0.988	-1.267

Figure 2.15.: Distribution of asymmetry measures for all flares that consist of more than one block derived with the baseline method (left) and half-clap method (right) for FSRQs (shaded red), BLLs (dotted blue), and UCs (gray contour). The table underneath shows the statistical moments of each distribution.

zero (i.e. the boxieness of the distribution) proves that all asymmetry measure values occur to a comparable degree.

Just as in Sec. 2.1.1, this analysis is also performed for each class of the monthly sample separately. The resulting distribution of asymmetry measures of all flares consisting of more than one block from all sources in the corresponding class is shown in Fig. 2.15. The baseline method is on the left and the half-clap method is on the right. FSRQs are shaded red, BLLs dotted blue, and UCs contoured gray. The table underneath contains the statistical moments (see Eq. 2.1.1) of each distribution. Each of these values is comparable with the value for the whole sample (see Fig. 2.14). Based on a KS test, it cannot be rejected that either combination of classes (from the same method) is drawn from the same distribution. Neither class shows a tendency for asymmetry or symmetry in their flux variations, based on these two methods. The distinction by blazar class does not yield a difference of rise and decay behavior of the corresponding flares.

2.4. Daily Flares

In the same manner, both flare finding methods were applied to the daily sample. 160 flares are obtained with the baseline methods, while the half-clap method yields 346 flares for the eight sources. Out of these, 66 and 171 are single-block flares, respectively. The discrepancy between the two methods is due to the fact that there are several flux variations that lie under the corresponding mean of the light curve. The daily light curves are plotted along with their monthly counter part in Figs. A.1-A.8 of the Appendix where the flares of each method are highlighted. Since the sampling of the daily sample is very dense, it is recommended to look at these light curves in the online version of this thesis[17]. Characteristics of the two flare finding algorithms are discussed in Sec. 3.2.1. The asymmetry measure for all flares of the eight daily sources is calculated according to Eq. 2.2.1. The resulting distributions $\tilde{A}_{d,bl}$ and $\tilde{A}_{d,hc}$ are shown in Fig. 2.16 on the left and right for the baseline and half-clap method, respectively. While the purpose of this work is to perform a sample study, discussion of individual flares and sources can be found elsewhere[18]

Also in this sample, there is an excess of $A = 0$, indicating a large proportion of symmetric flares. Exclusion of the single-block flares results in the distributions $A_{d,bl}$ and $A_{d,hc}$ shown in Fig. 2.16. This shows that also in this case the excess of symmetric flares is due to the ones of which the structure was not resolved detailed enough (see Sec. 3.2. Hence, the single-block flares are neglected for this sample as well. The table in that figure lists the values of the statistical moments of the distribution of asymmetry measures for each method excluding single-block flares.

In analogy to the results of the monthly sample, there is no tendency for asymmetry nor symmetry of the flux variations of the daily light curves based on the methods applied here.

---

[17] Thesis archive of the astronomy department Wuerzburg: https://www.physik.uni-wuerzburg.de/astro/archiv/theses-archive/, 24.05.2020

[18] Information on individual sources or particular flaring events can be found on the simbad website of the corresponding source, which is linked in Tab. 1.1.
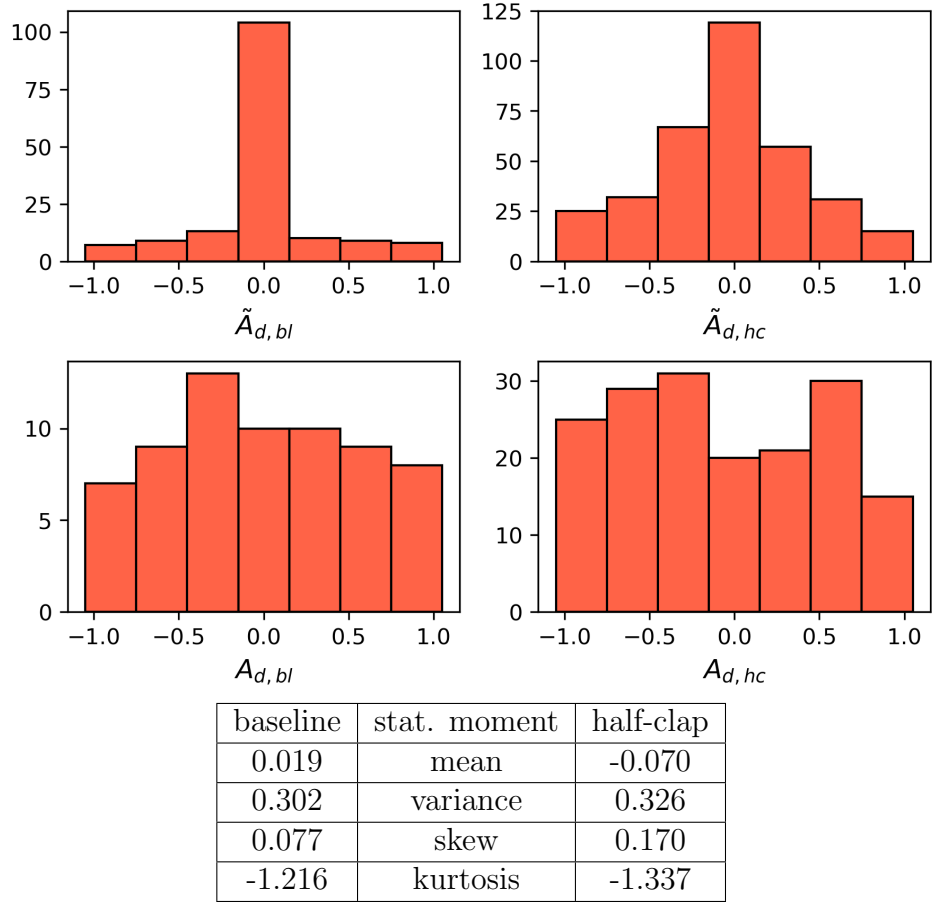
baseline	stat. moment	half-clap
0.019	mean	-0.070
0.302	variance	0.326
0.077	skew	0.170
-1.216	kurtosis	-1.337

Figure 2.16.: Distribution of asymmetry measures for all detected flares of the eight sources in the daily sample derived with the baseline method (top left) and half-clap method (top right). As in the monthly sample (Fig. 2.14), both methods yield each kind of asymmetry to a comparable degree when excluding single-block flares (bottom), as shown by the statistical moments of these distributions listed in the table.

3. Discussion

The focus of the first section of the discussion is on the flux distributions of the monthly and daily sample. Along with examples of the literature it is shown that the interpretation of flux distributions can basely lead to physical conclusions. In the second part, the two flare finding methods and their results are compared. Both yield symmetric and either kind of asymmetric flare shapes. The large fraction of single-block flares in both samples indicates that the true shape of the variations is not represented detailed enough. It is shown that this could be due to too wide time binning which suggests that the gamma-ray flux variations of blazars take place on sub-daily timescales. This is in accordance with other studies that detect gamma-ray flux variability on sub-hour timescales.

3.1. Interpretation of Flux Distributions

3.1.1. Monthly flux distributions

The illustration of the distribution of all monthly, scaled log-fluxes in Fig. 2.1 clearly shows that the raw light curves contain a distinct distribution of background/noise measurements, which has to be filtered. Requiring TS > 4 for each flux bin and that at least 50% of the flux bins in a light curve have to fulfill this criteria leaves 335 sources of the initial monthly sample that are suitable for the analysis. All of these sources are assumed to be blazars (see Sec. 2.1.1).

The statistical moments for each source of the monthly sample are as computed according to Eq. 2.1.1 in Sec. 2.1.1. Comparing the distribution of statistical moments according to source classes yields the following. Since the measured fluxes are distance dependent, comparison of the means is not meaningful here. It merely shows that each source class is measured to a comparable level. The variance, by contrast, is systemati-

cally larger for FSRQs than for BLLs. This means the former are more variable than the latter as has been found in previous studies (Ackermann et al., 2015). The hypothesis that the variance distributions of the two source classes are from the same distribution can be rejected with a p-value of 5% based on a KS test. For the skewness and kurtosis, this hypothesis can not be ruled out indicating that there is no significant distinction between the source classes based on the third or fourth statistical moment.

3.1.2. Daily flux distributions

In the daily sample, four sources have already been studied along with two other bright FSRQs in Meyer et al. (2019) (see Tab. 1.1). In Sec. 3 of that paper the authors present the flux distributions of weekly binned flux values (as opposed to daily binned flux values analyzed here).

The color coding in the daily histograms (see Figs. 2.3-2.10) aims to visualize stationarity. If each color spreads about the same mean value, the distribution can be assumed to be stationary. Another indication of stationarity is given by the time invariance of the mean value, which is estimated through computing segment-mean values and their variance. This only holds for the considered observation period (3600 days), while a longer or shorter stretch of the same data could yield different results. Furthermore, the considered steps of color coding and the size of segments for estimation of time invariance could change this conclusion. Bearing this in mind, the histograms suggest that three sources (OJ 287, PKS 0208-512, and PKS 2155-304) are stationary, two sources (3C 279 and 3C 454.3) appear non-stationary, and three are intermediate (PKS 0235+164, PKS 1510-089, and 3C 273).

The consideration of stationarity of the observed flux distributions is helpful for theoretical models about the production of the gamma-radiation. While it is very difficult to claim general (non-)stationarity by the observed data, these findings can be used to check blazar emission models. These need to be able to explain the rather stationary case such as PKS 2155-304 just as well as the rather non-stationary examples like 3C 454.3. Consistency checks should be conducted with various time scales and sizes of segments.

The color coded histograms have proven to lead to the same conclusions concerning this stationarity as the time invariance of the mean does. In fact, the colors visualize

the mean over time in the histogram. One can thereby represent more information than through an ordinary histogram. Nevertheless, firm conclusions on the stationarity of a distribution remain difficult.

3.1.3. Previous works on gamma-ray flux distributions

In Meyer et al. (2019) it is stated that the flux distributions of the considered FSRQs are well described by a log-normal or broken power law fit. The latter is motivated by self-organized criticality which is commonly observed in solar-flares as well (Aschwanden et al., 2016). The former is said to be often observed in the gamma-regime as discussed based on previous examples from the literature in the following.

The the most recent catalog of *Fermi*-detected AGN (4LAC The Fermi-LAT collaboration, 2019) does not provide flux distributions of individual sources. The previous version, 3LAC (Ackermann et al., 2015), covers 414 FSRQs and 604 BLLs from four years of observation. The flux values are binned monthly, resulting in 48 values that are then fitted with a log-normal function for each source. The motivation for this procedure, again, is based on the fact that a log-normal "has commonly been used to model blazar flux distributions." It is further stated that this "provides reasonable fits for most sources of [the] large sample."

Similarly, Shah et al. (2018) select data of 19 FSRQs and 19 BLLs covering about 100 months. They compare normal to log-normal fits for each flux distribution choosing adaptive binning for the histograms. This results in five bins for each distribution. For almost all sources the log-normal shape is preferred.

Tluczykont et al. (2010) is referred to as one of the common examples for log-normality in blazars. They analyze long-term VHE gamma-ray data, reporting that the flux distribution of the blazar Mrk 421 is best described with the sum of a Gaussian and a log-normal function.

The first reference of this kind with respect to blazars was Giebels & Degrange (2009). This research note reports a log-normal distribution of X-ray flux of the source BL Lacertae. All of the above publications refer to this work, particularly for posing implications on the observed log-normality, as shall be discussed later.

Source	Type	A_g	μ_g	σ_g	χ^2_R	N_{bins}	skew	kurtosis
OJ 287	BLL	89.038	-0.067	0.259	0.0713	26	0.079	-0.336
PKS 0208-512	BLL	178.707	-0.112	0.243	0.091	14	-0.133	-0.476
PKS 0235+164	BLL	99.153	-0.085	0.318	0.092	25	0.039	-0.503
PKS 1510-089	FSRQ	210.797	-0.169	0.379	0.190	42	0.087	0.170
PKS 2155-304	BLL	123.346	-0.090	0.241	0.158	30	-0.068	-0.433
3C 273	FSRQ	115.278	-0.071	0.321	0.263	42	0.324	0.447
3C 279	FSRQ	184.454	-0.229	0.384	0.240	42	0.385	0.797
3C 454.3	FSRQ	187.246	-0.177	0.408	0.204	51	0.002	0.854

Table 3.1.: Gaussian fit parameters and statistical moments of the logarithmic flux distribution of each source in the daily sample.

3.1.4. Log-normalcy of the daily flux distributions

In Sec. 2.1.2 of this work, the daily binned fluxes are fitted with a log-normal[1] as well. The obtained parameters are summarized in Tab. 3.1. The best-fit mean value μ_g of each log-normal distribution is systematically < 0. In the utilized scale, this means that the mean of the n logarithmic flux values f_i is always less than the logarithm of the mean of the flux values, i.e.

$$\frac{1}{n} \sum_i^n log(f_i) < log \left(\frac{1}{n} \sum_i^n f_i \right) \quad \Leftrightarrow \quad \prod_i^n f_i^{\frac{1}{n}} < \frac{1}{n} \sum_i^n f_i \tag{3.1.1}$$

which is mathematically expected for small values of f_i.

The variance σ_g is considered as a measure for the variability. Based on this, the statement that FSRQs are more variable than BLLs (Ackermann et al., 2015, also confirmed for monthly sample in the previous section) can be confirmed. The variability of PKS 0235+164 is closer to the order of the FSRQs than to the BLLs. As noted in the beginning, its classification as BLL is from simbad[2], while the *Fermi*-classification is uncertain.

The log-normal distribution seems to be an accurate representation of the daily flux histograms of OJ 287, PKS 0235+164, PKS 2155-304, and 3C 279. Besides the unnatu-

---

[1] Implemented by fitting a normal to the logarithmic flux distribution
[2] PKS 0235+164: http://simbad.u-strasbg.fr/simbad/sim-id?Ident=PKS+0235%2B164, 24.05.2020

rally high bin (see Sec. 2.1.2), also 3C 454.3 appears to be represented by the fit. Despite the fact that the sources PKS 0208-512, PKS 1510-089, and 3C 273 show an overall similar shape, the best fit appears off. As already mentioned in Sec. 2.1.2, this is due to the bias towards smaller bins given the Poissonian error estimation of the histogram bins.

Rather than judging the fits by eye, a chi-squared test allows to quantify the goodness of the fit. The fact that systematically $\chi^2_R < 1$ indicates that the errors were estimated too large. A quantitative comparison to the above studies is difficult because most of them do not note how (if at all) the flux uncertainty is propagated to the error of the histogram bins. Only Meyer et al. (2019) explain a way to propagate the error, which shall be applied in future work.

Of those who consider larger samples, only Shah et al. (2018) present values of χ^2_R for all of their log-normal fits. They range from 0.13 to 3.88 with most of the values < 1. This indicates that the errors were overestimated in most cases in this study as well. Since the procedure for computing this error is not elaborated, comparison is not possible.

Another way to analyze the log-normalcy of a distribution is to compare the statistical moments of the logarithmic values to a Gaussian. The skewness and kurtosis (Eq. 2.1.1) of the logarithmic flux values are listed in Tab. 3.1. The skewness, as introduced in the beginning of Sec. 2, quantifies the symmetry of a data sample. For symmetric distributions, such as a Gaussian or a centered box function, the skewness is zero. A negative (left) or positive (right) skew is denoted by the corresponding sign and is due to a large tail of low or high values. An absolute value of skewness > 1 is interpreted to indicate high skew (asymmetry). Any skew value between -0.5 and 0.5 is considered to have approximately no skew. All of the daily data sets analyzed in this study fall in the latter range and are therefore assumed to be symmetric. The sources 3C 273 and 3C 279 show the highest skew, with 0.324 and 0.385, as can be verified in the histograms (see Figs. 2.8 and 2.9). This means, based on the skewness, the distribution of logarithmic fluxes are similar to a Gaussian, which accounts for the log-normalcy of these fluxes.

The values for the kurtosis range from -0.503 to 0.854, which scatters about zero, as expected for a Gaussian.

Based on the results of this work, it is not feasible to draw firm conclusions on the log-normalcy of the daily flux distributions. The previous studies listed above do state that a log-normal behavior is possible. Yet it is often missed to point out how vague the

procedure is.

3.1.5. Implications of log-normal flux distributions

Assume that the distribution of the observed fluxes does follow the log-normal function. All of the above publications refer to Giebels & Degrange (2009) for physical interpretation. These authors first observed a log-normal flux distribution of a blazar, specifically X-ray flux of BL Lacertae. It is mistakenly claimed that such a log-normal flux distribution is indicating multiplicative physical processes as first raised by Uttley et al. (2005). In many papers referring to Giebels & Degrange (2009), the claim for multiplicativity and the physical explanation have been adapted. The theoretical model consulted is often based on variations in the accretion flow as proposed by Lyubarskii (1997). The author illustrates, that fluctuations in the (outer) accretion-disk propagate inwards and affect fluctuations in the accretion rate. This directly translates to flicker-noise type fluctuations in luminosity for the instance of X-ray binaries. Starting with Giebels & Degrange (2009), AGN-scientists adapt this explanation for gamma-rays by assuming that this behavior is, in a not further defined way, inherited by the jet. This does provide a possible explanation for the flicker-noise observed in gamma-rays. Alternatively, the fractal break-up of large coronal flares (magnetic reconnection) and further examples, for instance, listed in Rieger (2019) resemble a multiplicative process that could be present in blazars and accreting systems in general.

Often it is further mistaken that a linear relation between flux variability and the flux itself indicate multiplicative processes. This relation is given by the standard deviation and the mean of flux in a small interval of the light curve. Uttley & McHardy (2001) claim that such a linear relation is a common feature for compact accreting systems and suggest that the underlying physical process has to be multiplicative such as the ones mentioned above.

Scargle (2020) disproves conclusions like this. While log-normality and the rms-flux relation can be consequences of such models, it is wrong to assume that these properties necessarily trace back to only these models or multiplicativity in general, as is but to frequent in recent works (e.g. Rieger, 2019). Log-normality can also be obtained by an

additive process, specifically an auto regressive process, e.g.

$$X(n) = \sum_k c_k R(n-k) + D(n) \qquad \rightarrow \text{ log-normal } \textbf{or} \text{ normal} \qquad (3.1.2)$$

(from Scargle, 2020) where $R_k(n)$ represents an uncorrelated random process, c_k is a set of constant moving average coefficients, and $D(n)$ is a deterministic process (in practice often a constant). This disproves that log-normalcy can be asserted to certain processes underlying an astrophysical system. That is further affirmed by Burd et al. (in prep.), who found the best parameters for describing real light curves with an additive auto regressive process[3]. The authors show that the artificial light curves obtained from these parameters produce flux distributions of various kinds.

To conclude, a histogram appears to be a very simple first step to conduct when investigating data. However, as soon as further interpretation such as fits to this distribution are applied, careful consideration is necessary. This includes a comprehensive way to determine the binning, a reasonable error estimation for the height of the bins, and a measure for the goodness of possible fits. Without such quantitative measures, one can merely state phrases along the line "the fit describes the data well." Given that adequate procedures results in evidence for log-normality of the flux distribution, multiplicative processes such as variations in the accretion flow or flaring due to magnetic reconnection are possible explanations. Notwithstanding the lack of this evidence, it is not possible to rule out simpler, additive models that produce log-normal flux distributions (Burd et al., in prep. and Scargle, 2020). Thus, it is not possible to draw conclusions about the underlying physical process (either additive, multiplicative or a mixture of both) based on the distribution of observed fluxes.

3.2. Characterization of flux variations in light curves

The second approach to constrain the emission processes responsible for the high energy emission of blazars is to study flux variations by defining flares with the baseline and the half-clap method. These are discussed and compared in the following with a focus

---

[3] Ornstein-Uhlenbeck process

on time asymmetry. The results obtained in this analysis suggest that the variability of the high-energy emission of BLLs is not different to the one of FSRQs. Explanations for the various observed flare asymmetries are discussed.

3.2.1. Comparison of baseline and half-clap method

The two flare finding methods build up on the Bayesian Block algorithm. For the baseline method, one assumption is sufficient to objectively characterize flares in all the light curves: determining the value of the baseline that defines the onset and offset of a flare. As can be seen in Figs. A.1-A.8, this does capture the strong flux variations very well. Variations under the baseline flux are not considered. A severe problem is flux variations taking place above the baseline as a whole. These cases are interpreted as one flare consisting of several fluctuations with the peak at the highest block. For Fig. A.4 for instance, the fluctuations after MJD 57000 are interpreted as one flare. Moreover, it is not certain whether the sources can be assumed to have the same baseline throughout the whole observation period (i.e. assumed to be stationary, see Sec. 3.1). 3C 454.3, for instance, shows flux variations at different levels. The variation at MJD 55500 in Fig. A.8 is interpreted as one flare with this method, while this could also be several flux variations at a higher level than on other stretches of the light curve.

The assumption for the half-clap method is that the light curve consists of a quiescent state superposed with flaring events from the blazar. The rising and falling behavior of the flare can be extrapolated by clapping the adjacent block onto the block representing the quiescent level (generally long, low flux blocks). In case where two flares overlap, the middle of the block in between is considered as change point. The advantage of this method is that every local maxima of a flux variation is considered as its own flare. Naturally this affects that the half-clap method detects more flares than the baseline method. A disadvantage is that sometimes the blocks considered as quiescent state show minor variations themselves, which unnaturally stretches the extent of the flare. An example for this is the daily light curve of PKS 2155-304 in Fig. A.5. Due to the minor variation before MJD 56000, the rapid flare at MJD 56800 gets stretched.

A solution could be to combine the methods e.g. approach the concept with the half-clap method and only consider variations above a certain baseline limit. The HOP algorithm, introduced in Meyer et al. (2019), represents such a combination. The au-

thors further do not base their flares on the Bayesian blocks but fit every peak with an exponential rise and decay. Such improvements are going to be investigated in future work.

Based on the two methods applied here, the following features are systematically expected:

- Long half-clap flares due to variations in the quiescent state (e.g. Fig. A.5 at MJD 56000)

- Long baseline flares due to multiple peaks and/or a high state of the source (e.g. Fig. A.8 at MJD 55500)

- For a light curve with with many variations (daily), more half-clap than baseline flares are expected (multiple peaks)

- For a light curve with with little variations (monthly), it can happen that there are comparably few or no half-clap flares due to the conservative conditions applied here (e.g. start and end of light curve in Fig. A.2)

- Single block flares result in symmetric flares and cause a bias for $A = 0$

- Variations consisting of two Bayesian blocks with one data bin each cause a bias towards $A = \pm 0.5$ (e.g. Fig. A.4 at MJD 55100)

To take into account the former two features, a length filter or an upper limit for the asymmetry measure could be applied. This would, however, impose more assumptions on the expected shape of the variations. Since the goal of this work is to objectively study this shape, no such filter is invoked.

The remaining points are in accordance with the results presented in Sec. 2.3 and 2.4 (see Figs. 2.14 and 2.16). It is evident that there is a strong bias towards $A = 0$ which requires filtering for single block flares in both methods (see Sec. 3.2.4). The mentioned bias towards $A = \pm 0.5$ is neglected here.

3.2.2. Comparison of BLL vs FSRQ flares

The monthly asymmetry measures based on the flares consisting of more than one block are separated by source class in Fig. 2.15. The asymmetry measure for all classes are

very similar. Based on a KS test it can not be ruled out that the asymmetry measures of each class are drawn from the same distribution. That means, based on the findings of this work, the observed flux variations in FSRQs and BLLs are not distinguishable.

3.2.3. Meaning of the asymmetry measure

Connecting observed flux variations to their emission processes generally bears difficulties. After their initial emission, the photons could be absorbed and re-emitted. Additionally, the same process could take place at several locations at the same time and/or be superimposed with other processes (of different energies). Afterwards, the photon has to travel to and be measured by the detector. All of these steps correspond to convolutions that are important for such considerations.

Bearing these factors in mind, this work aims to give a systematic approach to characterize the shape of flux variations in term of defining flares with the baseline and the half-clap method. The detected flares in the eight daily and 2287 monthly binned *Fermi* light curves are then analyzed with respect to their asymmetry measure (Eq. 2.2.1). Based on this, three general types of flares can be distinguished, as discussed in the following. For now, it is assumed that each individual flare is due to a single event of a certain emission process. Later on, also superposition of the same process at different locations is considered. Taking into account more of the above difficulties lies beyond the scope of this thesis.

Negative asymmetry

A negative asymmetry measure or $A \to -1$ equals $t_r < t_d$. One exemplary case for this is a fast rise with an exponential decay (so-called FRED), as shown on the left of Fig. 2.13.
Flare shapes like these are for instance observed for supernovae of type 1a[4]. As pointed out in Hillebrandt et al. (2013), the uniformity of the SN1a luminosity variation is used as standard candle. It allows to determine the Hubble constant, which is topic of recent

---

[4] Thermonuclear explosion of a white dwarf that accretes matter from another companion. The white dwarf is teared apart by a shock wave forming a so-called ejecta. The heavier elements created via nuclear fusion emit photons through radioactive decay (^{56}Ni $\to$ ^{56}Co $\to$ ^{56}Fe).
https://d-nb.info/999423169/34, 18.05.2020

research in cosmology (Dhawan et al., 2018).

With respect to AGN and blazars, a negatively asymmetric high-energy flare shape could occur if a moving shock collides with a stationary formation associated with the core of the jet (Sokolov et al., 2004). Such collisions cause a forward and reverse shock[5] that are assumed to produce separate flares through both synchrotron and synchrotron self Compton emission. The plasma frame light curve obtained in such a scenario is shown in Fig. 3.1 for 10^{12} Hz (bottom) and 10^{17} Hz (top). Sokolov et al. (2004) point out that each flare is composed by the superposition of reverse and forward (shock) flare such as depicted. Within the frame of their assumptions, it is concluded that colliding shocks can explain symmetric or negatively asymmetric flares while flares with positive asymmetry (a broad, weak outburst followed by a spike of emission at the end) are not possible.

The shock-in-jet model was first suggested by Marscher & Gear (1985) in order to explained a millimeter-to-infrared flare[6] of 3C 273 (see Wagner, 2018, for an illustration). This theory is also evoked to explain the characteristic radio knots seen in the VLBI observations of AGN jets (Jorstad et al., 2017). The moving shock described in Sokolov et al. (2004) could cause the ejection of a new radio component like that by colliding with the stationary feature (Jorstad et al., 2001). It is stressed that this scenario is different to the collision of two moving shocks downstream of the stationary component associated with the jet core.

Symmetry

If the rise time is of the same order as the decay time, the asymmetry measure gets close to zero $A \rightarrow 0$. As pointed out previously, a large portion of the flares detected here are exactly zero. This is the case for single-block flares, which will be discussed separately in Sec. 3.2.4. This section focuses on flares that are traced with more than one block but still show a symmetric shape such as shown in the middle of Fig. 2.13.

One possible explanation for symmetric flares is streams of particles (electrons) within the plasma producing inverse Compton emission while they cross the line of sight. Giannios et al. (2009) propose a Poynting flux-dominated jet where magnetic reconnection

---

[5] Forward and reverse in the plasma frame; both forward in the rest frame of the AGN.

[6] At that time the theory of a uniformly expanding source was favored to explain blazar variability.
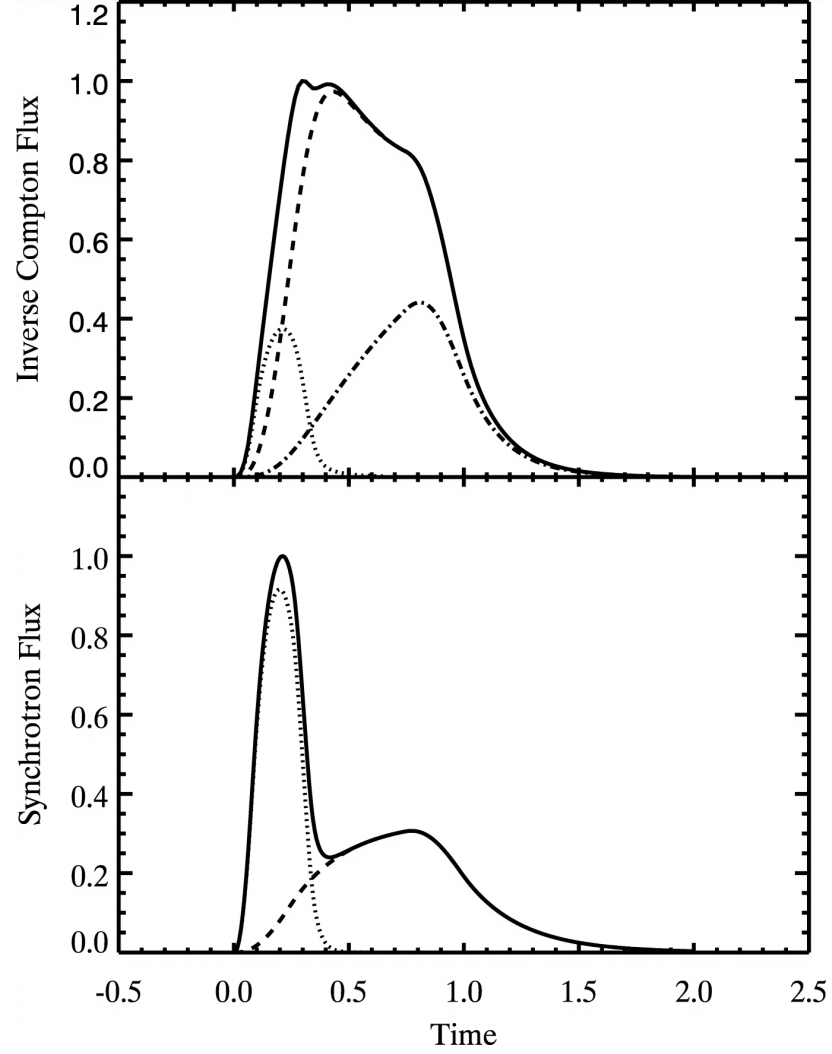
Figure 3.1.: Simulated forward, reverse, and overall light curves for 10^{17} Hz (inverse Compton; top) and 10^{12} Hz (synchrotron flux; bottom) in the plasma rest frame, depicted with dotted, dashed, and solid line respectively. The time axis is normalized to the apparent crossing time. Figure adapted from Sokolov et al. (2004)

takes place. Material that flows out from the reconnection region can have very high Lorentz factors. This can produce flares via synchrotron self Compton emission. Considering the structure of the dissipation region producing the so-called jets in a jet[7] it is conceivable that these have an angular velocity with respect to the observer. For a constant angular velocity, the flux observed from such an event would steadily increase up to the maximum when the mini jet is pointed towards us[8]. The following decrease would have the same form as the increase resulting in a symmetric flare.

Ghisellini et al. (2009) suggest that electrons could stream along magnetic field lines in the jet causing beams. If these lines change their orientation, the beam could cross the line of sight, causing a symmetric flare in analogy to the scenario above.

Both of these theories refer mainly to TeV photons but could also be possible in the *Fermi*-regime.

Positive asymmetry

If $t_r > t_d$ the asymmetry measure is positive or $A \to +1$. Physical models for emission processes that cause a flux variation of this form are challenging. External influences such as obscuration could pose a possible explanation for a rapid decline of flux like that.

Another explanation could be that the observed variability is, in fact, composed of a superposition of single events. Any form of asymmetric flare could be composed of several symmetric flux variations. Comparing the shape of single flares to overlapping ones, as planned for future work, could help to give insights on that.

3.2.4. Comparison of monthly and daily flares

The flares of the monthly and daily sample show similar distributions of asymmetry measures. In both Fig. 2.14 and Fig. 2.16 every type of asymmetry occurs. Flares with an asymmetry of almost zero are the most abundant in both methods for both samples. Filtering single-block flares shows that these are the reason for that bias. A single block is not sufficient to represent the ground laying shape of the flux variation.

---

[7] See Fig. 1 in Giannios et al. (2009).
[8] Amplified due to Doppler boosting.

Consider for instance a negatively asymmetric[9], symmetric[10], or positively asymmetric[11] flux variation as shown on the left, middle, and right of the top row in Fig. 3.2. For simplicity[12], flux measurements of these artificial flare shapes are assumed to be the y-axis value of the corresponding shape at a certain time on the x-axis. A relative y-axis error of 10% is considered for measurement uncertainty. Sufficiently small steps on the x-axis ("time binning"), as shown in the middle row, allows to reproduce the initial flare shape with Bayesian blocks. However, if the steps on the x-axis ("time binning") are too large, each of the flare shapes results in a single block. Since this is due to the data binning, this effect remains regardless of the parameters of the Bayesian block algorithm. Thus, flares that consist of one block only are conservatively dismissed in this analysis.

In the monthly sample this affects about 50% of all flares in both methods. This is in accordance with the result that variability on the timescales of days is found in the corresponding daily light curves. An example for this is shown in Fig. 3.3. For the purpose of this work the different energy ranges for the monthly and daily sample ($E > 1\,\text{GeV}$ and $E > 100\,\text{MeV}$, respectively) are neglected[13]. The flux variations on timescales of days are integrated over 28 days and thereby washed out in monthly binning. When applying the Bayesian block algorithm, this often results in blocks containing one bin only (see top panel). In addition to that, it has to be taken into account that flares that consist of several blocks in the monthly sample usually trace back to distinct variations on smaller time scales. A clear instance for this is the flux behavior in the end of the light curves in Fig. 3.3. Further examples (Figs. A.1-A.8) prove that the flux fluctuations on monthly timescales are typically due to shorter variations traced in daily binning. It is therefore concluded that the gamma-ray variability of blazars takes place on smaller timescales than months.

In the daily sample, 21 and 37 out of 50 and 54 flares (42% and 68%) consist of one block only, for the baseline and half-clap method respectively. Following the previous paragraph, this could indicate that the gamma-ray variability of blazars takes place on sub-daily timescales.

---

[9] $F(t) = \left[exp\left(t/t_d\right)\right) + exp\left(-t/t_r\right)\right]^{-1}$, from Meyer et al. (2019) with $t_r = 5$; $t_d = 50$
[10] $F(t) = \sqrt{2\pi}\sigma exp\left(-x^2/2\sigma^2\right)$, Gaussian
[11] $F(t) = \left[exp\left(t/t_d\right)\right) + exp\left(-t/t_r\right)\right]^{-1}$, from Meyer et al. (2019) with $t_r = 50$; $t_d = 5$
[12] Flux bins of the *Fermi* light curves represent integrated flux over the given time bin.
[13] While details of the light curve can differ for different energy ranges within *Fermi*, the overall shape is fairly similar (see e.g. Ackermann et al., 2016).
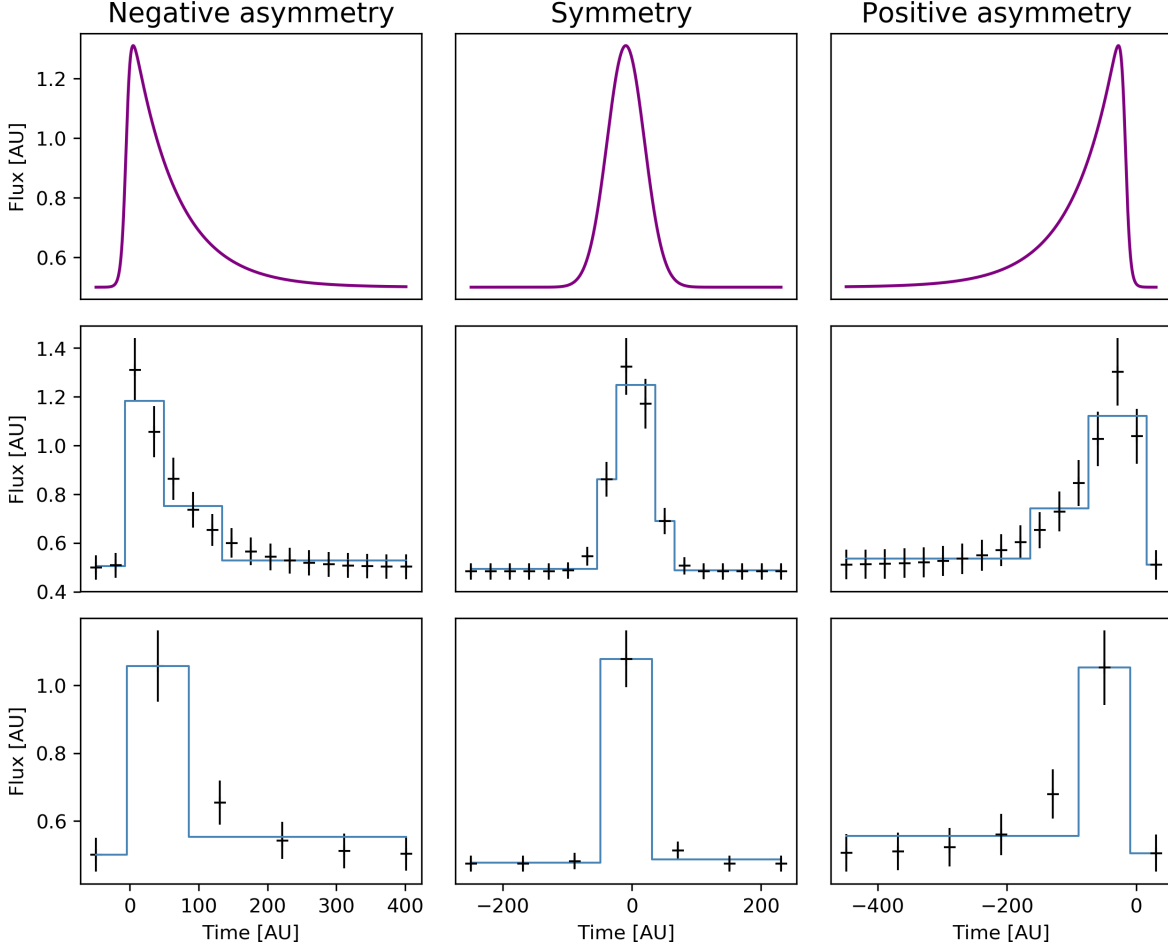
Figure 3.2.: The Bayesian block representation of flares is affected by the time binning of the data. The top row shows the "true" flare shape, while the middle and bottom row are "measurements" of these shapes at specific times with an uncertainty of 10%. While the binning in the middle panel is sufficiently small to represent the flare shape with Bayesian blocks, the binning in the bottom panel washes out the flare shape. This results in a single-block flare for every shape.
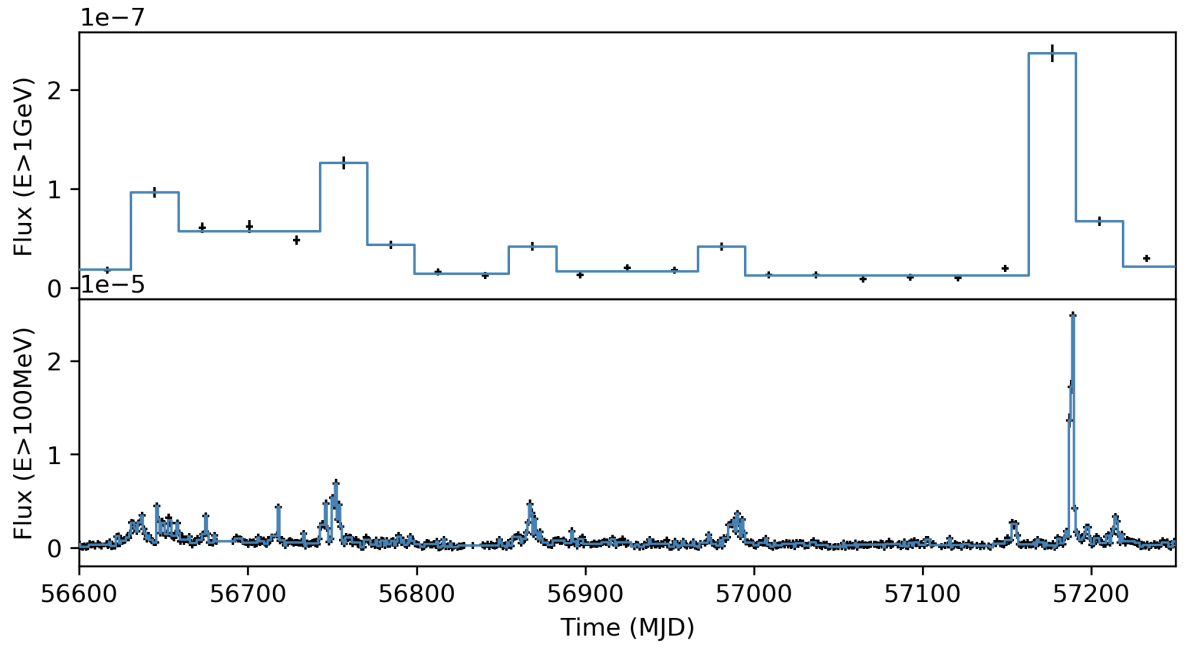
Figure 3.3.: Light curve of 3C 279 in monthly binning for energies over 1 GeV (top) and daily binning for energies over 100 MeV (bottom). Too wide of time binning can obscure the variability timescales of gamma-ray flux from blazars.

Ackermann et al. (2016) show with the instance of 3C 279 that high energy flux variations of blazars can take place on minute timescales. This extreme event was observed with *Fermi*-LAT on the 16th of June 2016 and is also visible in Fig. 3.3 at MJD 57189. The minute scale variability is washed out even in daily binning.

Meyer et al. (2019) reproduce this result and find minute scale variability for the blazar CTA 102 as well. The authors lay out that further FSRQs show flux variations on sub-hour time scales.

The limitation to resolve such timescales also at lower flux levels is that there must be enough photons to provide good statistics for the measurements. It is possible that this limitation precludes observation of the true gamma-ray variability timescales of blazars.

3.3. Conclusions and Outlook

Analysis of the statistical moments of the log-flux distributions from 335 monthly binned *Fermi*-sources according to their source type shows that the variance is significantly larger for FSRQs than for BLLs. The conclusion that FSRQs are more variable than BLLs is in accordance with previous results. Neither of the other three statistical moments shows significant difference between the source classes.

This result can be reproduced for eight of these sources in daily binning as well. Moreover, Color coded histograms on the one hand and time invariance of the mean on the other hand indicate that the flux distributions of these blazars range from non-stationarity to stationary (with respect to observation period). Possible emission models are required to reproduce either of these cases.

The log-normalcy of the daily flux distributions is ambiguous. Generally, the shape of a flux distribution is not suited to draw firm conclusions on (the multiplicativity or additivity of) the ground laying physical processes.

The two approaches to define flares (baseline and half-clap method) yield an excess of symmetric flares in the monthly as well as in the daily binning. In both samples, this is due to single-block flares. Since these result from variations that are not resolved detailed enough, this is a hint for too wide time binning. Either kind of (a)symmetry and symmetry occur to a comparable degree in both samples, when excluding such single-block flares. Consequently, it is not possible that one particular emission process (causing only

a positively/negatively asymmetric or a symmetric flare on the observed timescales) is responsible for the individual flux variations. It is very likely that the observed flux variations are a superposition of distinct events that take place on sub-daily timescales. Resolving such flux fluctuations is limited by the statistics of measuring gamma-photons with *Fermi*.

The following tasks are planned for future work:

- Constraining physical expectations from the emission models (i.e. what are the expected timescales?)

- Taking into account flux bins with low significance (TS) for the Bayesian block analysis (e.g. "upper limits")

- Propagating the uncertainty of the flux measurements to the histogram bins of the distribution (Meyer et al., 2019)

- Combining baseline and half-clap method to a more powerful flare finding algorithm.

- Applying these methods to more/other data such as the first GeV outburst of the radio-loud narros-line Seyfert 1 galaxy PKS 2004-447 (Gokus et al., in prep.) or to synthetic light curves produced with Ornstein-Uhlenbeck parameters derived from real *Fermi* measuremetns (Burd et al., in prep.)

- Combine this method with physical models (e.g. with exponential fits, see Garrappa et al., in prep.)

Bibliography

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

Ackermann M., Ajello M., Atwood W.B., et al., 2015, ApJ 810, 14

Ackermann M., Anantua R., Asano K., et al., 2016, ApJL 824, L20

Aschwanden M.J., Crosby N.B., Dimitropoulou M., et al., 2016, Space Sci. Rev. 198, 47

Atwood W., Albert A., Baldini L., et al., 2013, arXiv e-prints arXiv:1303.3514

Atwood W.B., Abdo A.A., Ackermann M., et al., 2009, ApJ 697, 1071

Blandford R.D., Payne D.G., 1982, MNRAS 199, 883

Blandford R.D., Znajek R.L., 1977, MNRAS 179, 433

Boettcher M., Harris D., Krawczynski H., 2012, Relativistic Jets from Active Galactic Nuclei, Wiley

Dhawan S., Jha S.W., Leibundgut B., 2018 609, A72

Eisenstein D.J., Hut P., 1998, ApJ 498, 137

Finke J., 2016, arXiv e-prints arXiv:1602.05965

Ghisellini G., Tavecchio F., Bodo G., Celotti A., 2009, MNRAS 393, L16

Ghisellini G., Tavecchio F., Foschini L., Ghirland a G., 2011, MNRAS 414, 2674

Giannios D., Uzdensky D.A., Begelman M.C., 2009, MNRAS 395, L29

Giebels B., Degrange B., 2009 503, 797

Hillebrandt W., Kromer M., Röpke F.K., Ruiter A.J., 2013, Frontiers of Physics 8, 116

Hovatta T., Petropoulou M., Richards J.L., et al., 2015, MNRAS 448, 3121

IceCube CollaborationAartsen M.G., Ackermann M., et al., 2018, Science 361, eaat1378

Jorstad S.G., Marscher A.P., Mattox J.R., et al., 2001, The Astrophysical Journal 556, 738

Jorstad S.G., Marscher A.P., Morozova D.A., et al., 2017, ApJ 846, 98

Kellermann K.I., Sramek R., Schmidt M., et al., 1989, AJ 98, 1195

Knuth K.H., 2006, arXiv e-prints physics/0605197

Kovačević M., Chiaro G., Cutini S., Tosti G., 2020, MNRAS 493, 1926

Liodakis I., Blinov D., 2019, MNRAS 486, 3415

Lorenz E., MAGIC Collaboration 2004 48, 339

Lyubarskii Y.E., 1997, MNRAS 292, 679

Mannheim K., 1993, A&A 269, 67

Marscher A.P., Gear W.K., 1985, ApJ 298, 114

Mattox J.R., Bertsch D.L., Chiang J., et al., 1996, ApJ 461, 396

Meyer M., Scargle J.D., Blandford R.D., 2019, ApJ 877, 39

Peñil P., Domínguez A., Buson S., et al., 2020, arXiv e-prints arXiv:2002.00805

Rieger F., 2019, Galaxies 7, 28

Rybicki G.B., Lightman A.P., 1986, Radiative Processes in Astrophysics

Scargle J.D., 2020, arXiv e-prints arXiv:2001.08314

Scargle J.D., Norris J.P., Jackson B., Chiang J., 2013, ApJ 764, 167

Schmitt J.L., 1968, Nat 218, 663

Shah Z., Mankuzhiyil N., Sinha A., et al., 2018, Research in Astronomy and Astrophysics 18, 141

Sokolov A., Marscher A.P., McHardy I.M., 2004, ApJ 613, 725

The Fermi-LAT collaboration 2019, arXiv e-prints arXiv:1905.10771

Tluczykont M., Bernardini E., Satalecka K., et al., 2010 524, A48

Urry C.M., 1999, Astroparticle Physics 11, 159

Urry C.M., Padovani P., 1995, PASP 107, 803

Uttley P., McHardy I.M., 2001, MNRAS 323, L26

Uttley P., McHardy I.M., Vaughan S., 2005, MNRAS 359, 345

Van Rossum G., Drake F.L., 2009, Python 3 Reference Manual, CreateSpace, Scotts Valley, CA

Vermeulen R.C., Ogle P.M., Tran H.D., et al., 1995, ApJL 452, L5

Wagner S.M., 2018, VLBI Jet Kinematics of Mrk 421 concerning the Radio Flare in 2012

A. Appendix

A.1. Bayesian Block Implementation

```python
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.colors as colors
import astropy.stats.bayesian_blocks as bblocks

def BB_4_LC(time, flux ,flux_error, gamma_value=None, p0_value=None):
        block_pbin=np.zeros(len(flux))
        if gamma_value is None and p0_value is None:
                edges = bblocks(t=time, x=flux, sigma=flux_error, fitness='measures')
        elif gamma_value is None:
                edges = bblocks(t=time, x=flux, sigma=flux_error, fitness='measures',
                        p0=p0_value)
        elif p0_value is None:
                edges = bblocks(t=time, x=flux, sigma=flux_error, fitness='measures',
                        gamma=gamma_value)
        else:
                edges = bblocks(t=time, x=flux, sigma=flux_error, fitness='measures',
                        gamma=gamma_value, p0=p0_value)

        if len(edges) <= 2:
                # print('This LC is constant, only one Bayesian Block found')
                return block_pbin, np.zeros(0), np.zeros(0), 0, edges
        else:
                #get edge_index to access other arrays
                edge_index = np.array([np.where(time >= edges[i])[0][0]
                        for i in range(len(edges))])
                #change last entry such that loop over [j:j+1] gives all BBs
                edge_index[-1] += 1

                #determine flux value (mean) and error (Gaussian propagation)
                block_val = np.zeros(len(edge_index)-1)
                block_val_error = np.zeros(len(edge_index)-1)
                for j in range(len(edge_index)-1):
                        block_val[j] = np.mean(flux[edge_index[j]:edge_index[j+1]])
                        block_val_error[j] = np.sqrt(np.sum(flux_error[edge_index[j]:
                                edge_index[j+1]]**2))/(edge_index[j+1]-edge_index[j])

                #create BB array corresponding to data
                for k in range(len(block_val)):
                        block_pbin[edge_index[k]:edge_index[k+1]] = block_val[k]

        return block_pbin, block_val, block_val_error, edge_index, edges
```

A.2. Flare Finding Baseline Method

```python
def FF_baseline (baseline, block_val, edges,
        LC_edges='neglect', talking=False):

        #handeling edges of LC:
        #LC_edges == 'add' -> start/ end is artificially added at start/end of LC
        #LC_edges == 'neglect' -> incomplete flares are conservatively neglected

        diff = np.diff(block_val)

        #all local peaks over baseline
        peak_time=[]
        peak_block=[]
        for i in range(1,len(diff)):
                #previous rising; this falling
                if diff[i-1]>0 and diff[i]<0:
                        if block_val[i]>baseline:
                                peak_time.append(edges[i] + (edges[i+1]-edges[i])/2)
                                peak_block.append(i)
        peak_time=np.array(peak_time)
        peak_block=np.array(peak_block)

        #all possible starts and ends (crossing baseline)
        start_time=[]   #starts with first block over baseline
        start_block=[] #left block under baseline
        end_time=[]     #ends with last block over baseline
        end_block=[]    #right block under baseline
        for i in range(len(block_val)-1):
                if block_val[i]<baseline and block_val[i+1]>baseline:
                        start_time.append(edges[i+1])
                        start_block.append(i)
                if block_val[i]>baseline and block_val[i+1]<baseline:
                        end_time.append(edges[i+1])
                        end_block.append(i+1)
        start_time=np.array(start_time)
        start_block=np.array(start_block)
        end_time=np.array(end_time)
        end_block=np.array(end_block)

        ## HANDELING EXCEPTIONS same as baseline method
        #at least one full flare has to be found (*)
        if len(peak_block)<1:
                if talking==True:
                        print('not variable enough, no peak found')
                return(np.array([0]), peak_block, start_time,
                        start_block, end_time, end_block,0)
        if len(start_block)<1 or len(end_block)<1:
                if talking==True:
                        print('not variable enough, missing start or end')
```

```python
            return(np.array([0]), peak_block, start_time,
                    start_block, end_time, end_block,0)


    #single end at end of LC is deleted
    if end_time[0]<peak_time[0]:
            end_block = np.delete(end_block,0)
            end_time = np.delete(end_time,0)
            if talking==True:
                    print('deleted single end in beginning of LC')
            if len(end_time)<1:
                    if talking==True:
                            print('this was the only end, not variable enough')
                    return(np.array([0]), peak_block, start_time,
                            start_block, end_time, end_block,0)
    #single start at beginning of LC is deleted
    if start_time[-1]>peak_time[-1]:
            start_block = np.delete(start_block,-1)
            start_time = np.delete(start_time,-1)
            if talking==True:
                    print('deleted single start in end of LC')
            if len(start_time)<1:
                    if talking==True:
                            print('this was the only start, not variable enough')
                    return(np.array([0]), peak_block, start_time,
                            start_block, end_time, end_block,0)


    #peak without start in beginning of LC
    #(there must be at least one other full flare bc of (*))
    if peak_time[0]<start_time[0]:
            #artificially add flare
            if LC_edges == 'add':
                    start_time = np.insert(start_time,0,edges[0])
                    start_block = np.insert(start_block,0,-1)
                    if talking==True:
                            print('inserted single start in beginning of LC')
            #conservatively dismiss first peak if there are multiple peaks
            while LC_edges == 'neglect' and end_time[0]>peak_time[1]:
                    peak_time = np.delete(peak_time,0)
                    peak_block = np.delete(peak_block,0)
                    if talking==True:
                            print('neglected 1st multiple peak in beginning of LC')
            #conservatively dismiss first peak and first end
            if LC_edges == 'neglect' and end_time[0]<peak_time[1]:
                    peak_time = np.delete(peak_time,0)
                    peak_block = np.delete(peak_block,0)
                    end_time = np.delete(end_time,0)
                    end_block = np.delete(end_block,0)
                    if talking==True:
```

```python
                                print('neglected peak + end in beginning of LC')
                        if len(peak_block)<1 or len(end_block)<1:
                                if talking==True:
                                        print('not variable enough')
                                return(np.array([0]), peak_block, start_time,
                                        start_block, end_time, end_block,0)


        #peak without end in end of LC
        #(there must be at least one other full flare bc of (*))
        if peak_time[-1]>end_time[-1]:
                #artificially add flare
                if LC_edges == 'add':
                        end_time = np.append(end_time,edges[-1])
                        end_block = np.append(end_block,len(block_val))
                        if talking==True:
                                print('inserted single end in end of LC')
                #conservatively dismiss last peak if there are multiple peaks
                while LC_edges == 'neglect' and start_time[-1]<peak_time[-2]:
                        peak_time = np.delete(peak_time,-1)
                        peak_block = np.delete(peak_block,-1)
                        if talking==True:
                                print('neglected last multiple peak in end of LC')
                #conservatively dismiss last peak and last start
                if LC_edges == 'neglect' and start_time[-1]>peak_time[-2]:
                        peak_time = np.delete(peak_time,-1)
                        peak_block = np.delete(peak_block,-1)
                        start_time = np.delete(start_time,-1)
                        start_block = np.delete(start_block,1)
                        if talking==True:
                                print('neglected peak + start in end of LC')


        ##single-peak mask to flag multiple peaks within start and end
        peak_mask = np.zeros(len(peak_block), dtype =bool)
        for x in range(len(end_block)):
                real_peak_block = start_block[x] +
                        np.argmax(block_val[start_block[x]:end_block[x]])
                mask_index = np.where(peak_block == real_peak_block)[0][0]
                peak_mask[mask_index] = True


        return(peak_time, peak_block, start_time,
                start_block, end_time, end_block, peak_mask)
```

A.3. Flare Finding Half Clap Method

```python
def FF_sarah (block_val, edges,
        method = 'halfclap', LC_edges = 'neglect', talking = False):

        # method = how is start and end defined
        # 'half' -> start and end are at the middle of change block
        # 'halfclap' -> min(middle and clap)
        # 'sharp' -> ignore change block, cut out rest

        diff = np.diff(block_val)
        peak_block = []
        peak_time = []
        start_block = []
        start_time = []
        end_block = []
        end_time = []

        #all local peaks and changes
        for i in range(1,len(diff)):
                #previous rising; this falling
                if diff[i-1]>0 and diff[i]<0:
                        peak_block.append(i)
                        peak_time.append(edges[i] + (edges[i+1]-edges[i])/2)
                #previous falling; this rising
                if diff[i-1]<0 and diff[i]>0:
                        half_block_time = (edges[i+1]-edges[i])/2
                        start_block.append(i)
                        end_block.append(i)
                        if method == 'half':
                                start_time.append(edges[i+1] - half_block_time)
                                end_time.append(edges[i] + half_block_time)
                        if method == 'halfclap':
                                start_time.append(edges[i+1] - np.minimum(
                                        half_block_time, edges[i+2]-edges[i+1]))
                                end_time.append(edges[i] + np.minimum(
                                        half_block_time, edges[i]-edges[i-1]))
                        if method == 'sharp':
                                start_time.append(edges[i+1])
                                end_time.apend(edges[i])

        ## HANDELING EXCEPTIONS same as baseline method
        #at least one full flare has to be found (*)
        if len(peak_block)<1:
                if talking==True:
                        print('not variable enough, no peak found')
                return(np.array([0]), peak_block, start_time,
                        start_block, end_time, end_block)
        if len(start_block)<1 or len(end_block)<1:
                if talking==True:
```

```python
                print('not variable enough, missing start or end')
        return(np.array([0]), peak_block, start_time,
                    start_block, end_time, end_block)


    #single end at end of LC is deleted
    if end_time[0]<peak_time[0]:
        end_block = np.delete(end_block,0)
        end_time = np.delete(end_time,0)
        if talking==True:
                print('deleted single end in beginning of LC')
        if len(end_time)<1:
                if talking==True:
                        print('this was the only end, not variable enough')
                return(np.array([0]), peak_block, start_time,
                        start_block, end_time, end_block)
    #single start at beginning of LC is deleted
    if start_time[-1]>peak_time[-1]:
        start_block = np.delete(start_block,-1)
        start_time = np.delete(start_time,-1)
        if talking==True:
                print('deleted single start in end of LC')
        if len(start_time)<1:
                if talking==True:
                        print('this was the only start, not variable enough')
                return(np.array([0]), peak_block, start_time,
                        start_block, end_time, end_block)


    #peak without start in beginning of LC
    #(there must be at least one other full flare bc of (*))
    if peak_time[0]<start_time[0]:
        #artificially add flare
        if LC_edges == 'add':
                start_time = np.insert(start_time,0,edges[0])
                start_block = np.insert(start_block,0,-1)
                if talking==True:
                        print('inserted single start in beginning of LC')
        #conservatively dismiss first peak if there are multiple peaks
        while LC_edges == 'neglect' and end_time[0]>peak_time[1]:
                peak_time = np.delete(peak_time,0)
                peak_block = np.delete(peak_block,0)
                if talking==True:
                        print('neglected 1st multiple peak in beginning of LC')
        #conservatively dismiss first peak and first end
        if LC_edges == 'neglect' and end_time[0]<peak_time[1]:
                peak_time = np.delete(peak_time,0)
                peak_block = np.delete(peak_block,0)
                end_time = np.delete(end_time,0)
                end_block = np.delete(end_block,0)
```

```python
                        if talking==True:
                                print('neglected peak + end in beginning of LC')
                        if len(peak_block)<1 or len(end_block)<1:
                                if talking==True:
                                        print('not variable enough')
                                return(np.array([0]), peak_block, start_time,
                                        start_block, end_time, end_block)


    #peak without end in end of LC
    #(there must be at least one other full flare bc of (*))
    if peak_time[-1]>end_time[-1]:
            #artificially add flare
            if LC_edges == 'add':
                    end_time = np.append(end_time,edges[-1])
                    end_block = np.append(end_block,len(block_val))
                    if talking==True:
                            print('inserted single end in end of LC')
            #conservatively dismiss last peak if there are multiple peaks
            while LC_edges == 'neglect' and start_time[-1]<peak_time[-2]:
                    peak_time = np.delete(peak_time,-1)
                    peak_block = np.delete(peak_block,-1)
                    if talking==True:
                            print('neglected last multiple peak in end of LC')
            #conservatively dismiss last peak and last start
            if LC_edges == 'neglect' and start_time[-1]>peak_time[-2]:
                    peak_time = np.delete(peak_time,-1)
                    peak_block = np.delete(peak_block,-1)
                    start_time = np.delete(start_time,-1)
                    start_block = np.delete(start_block,-1)
                    if talking==True:
                            print('neglected peak + start in end of LC')


    return(peak_time, peak_block, start_time, start_block, end_time, end_block)
```

A.4. Light curves and flares of the daily sample

The following plots show the light curves of the sources in the daily sample. The daily binned light curve is plotted underneath the corresponding monthly binned light curve along the same time axis in MJD. The data and errorbars are shown in a light blue, overlaid with the Bayesian blocks (obtained with default parameters) in blue.

The blocks that belong to flares defined with the baseline method are filled with turquoise. Flares detected with the half-clap method are alternately shaded in orange and orchid over the whole y-axis of the light curve since they are typically longer than the baseline flares (See Sec. 3.2.1 for discussion of the methods). Comparison of monthly and daily blocks shows that variations typically take place on smaller timescales than months.

A closer investigation of the crowded daily light curves is recommended in the online-version of this thesis[1].

---

[1] See: https://www.physik.uni-wuerzburg.de/astro/archiv/theses-archive/, 26.05.2020)
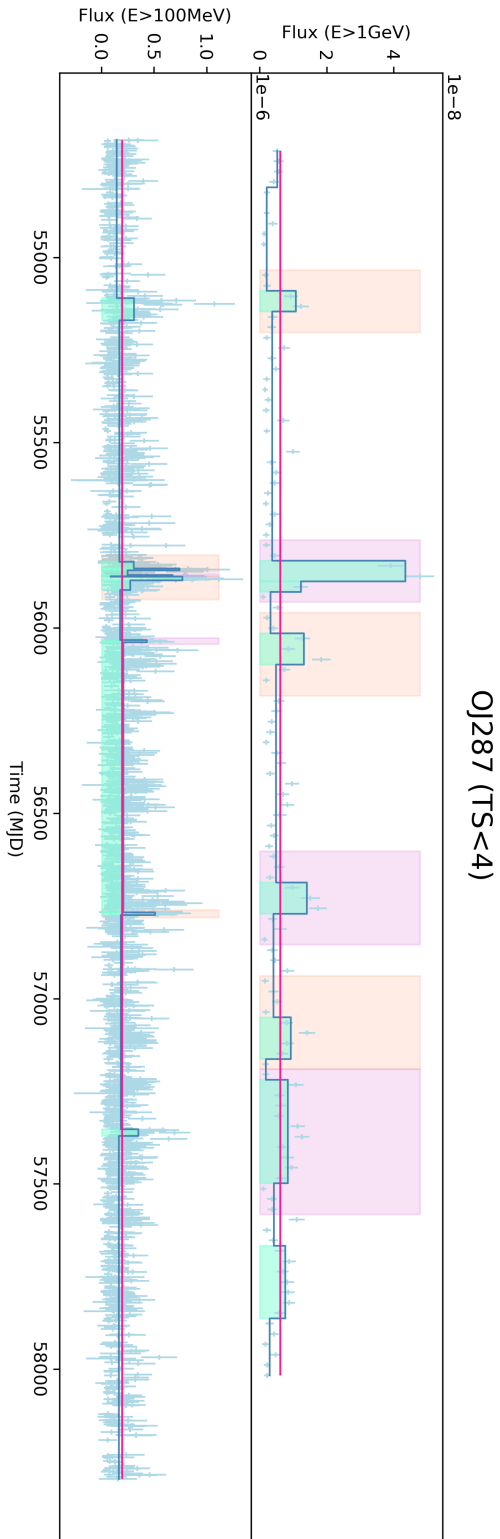
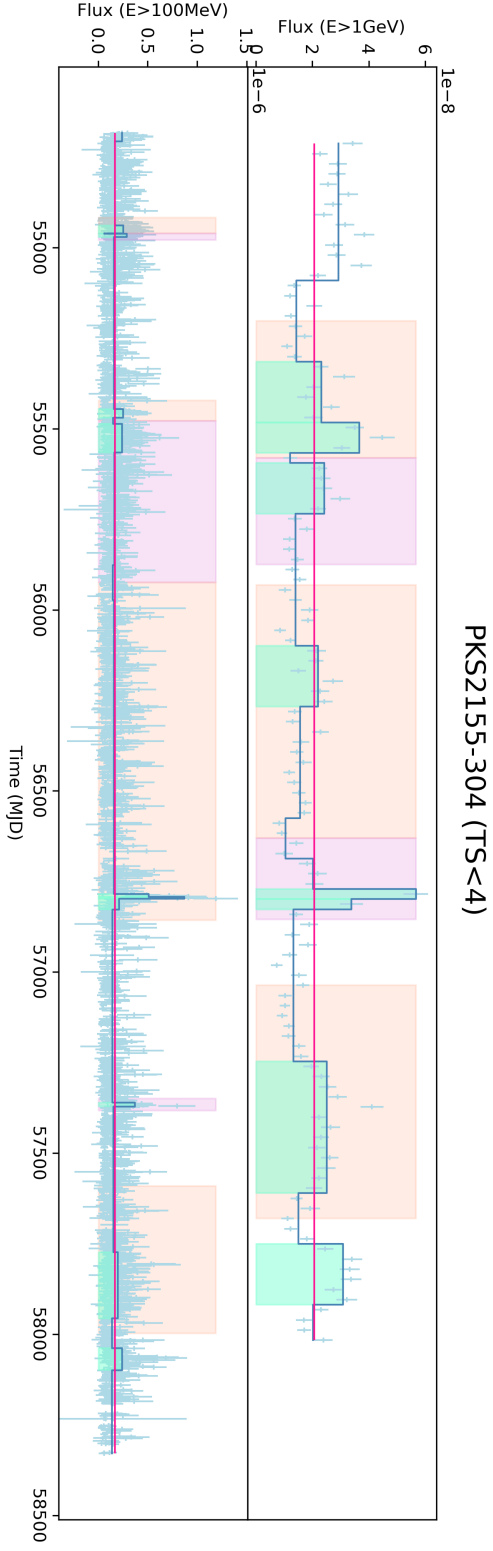**Figure A.1.:** Light curve with observed flux of OJ 287 in monthly (*top*) and daily (*bottom*) binning for energies over 1GeV and 100MeV, respectively. The baseline method defines Bayesian Blocks to belong to a flare (shaded turquoise) if the flux level exceeds the mean plotted in pink. The flares of the half-clap method (shaded in pink and orange) are defined at their start and end by the minimum of half the block length and extrapolating the slope onto the valley block. (See Sec. 2.2 for further information.)
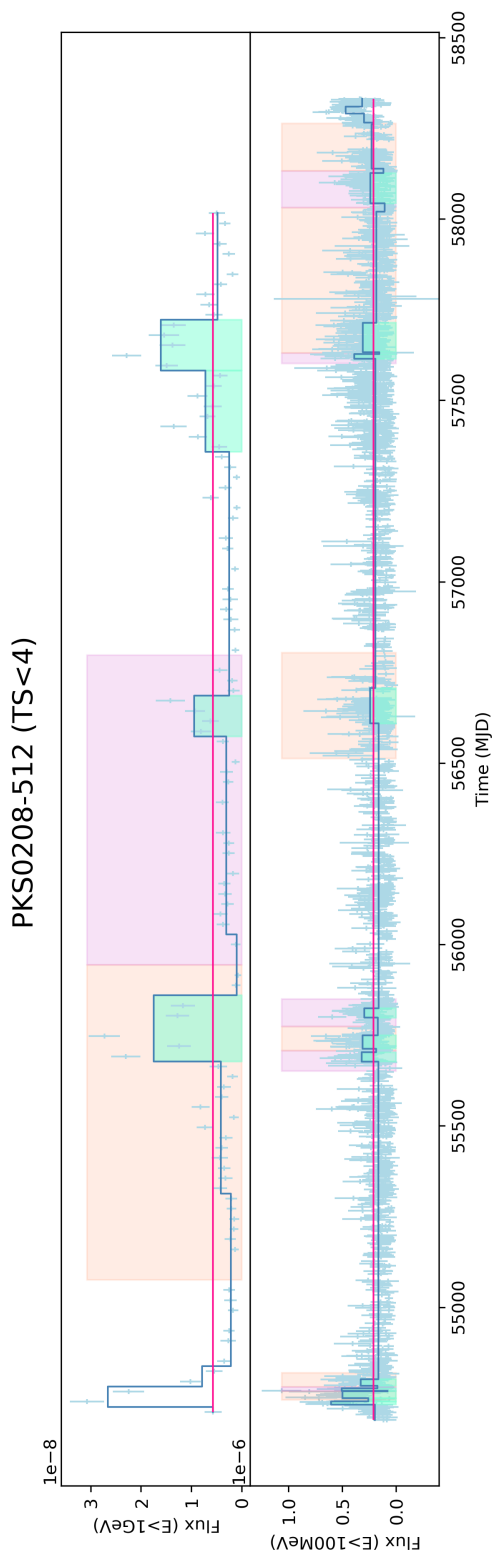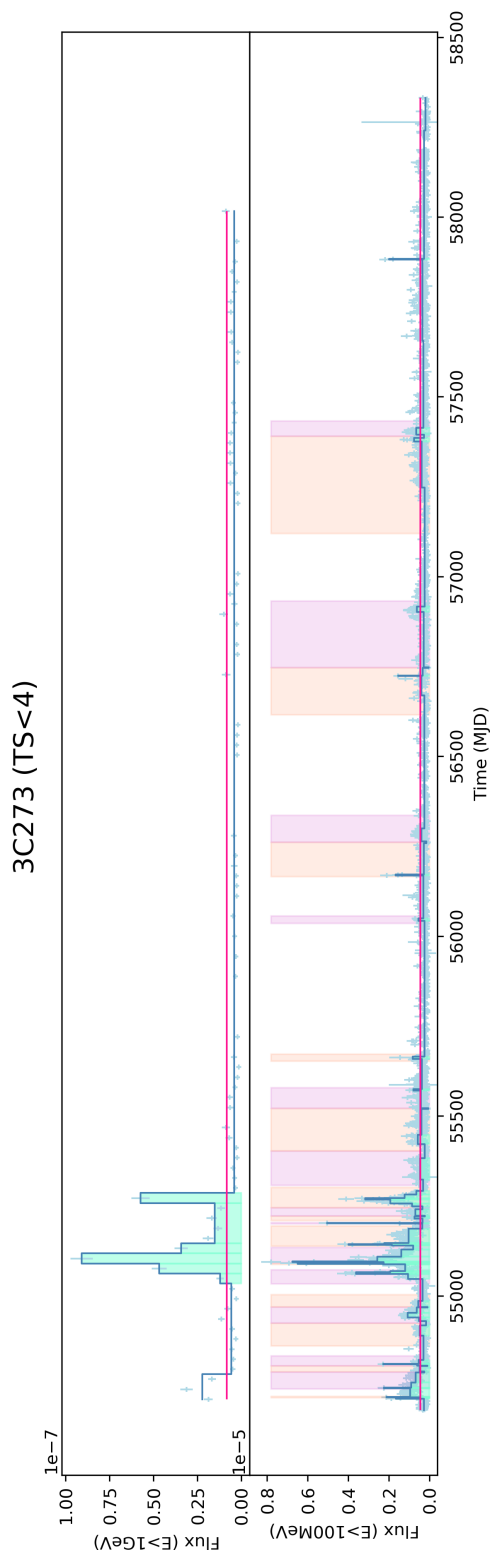
Figure A.2.:: Light curve with observed flux of PKS 0208-512 in monthly (*top*) and daily (*bottom*) binning for energies over 1GeV and 100MeV, respectively. The baseline method defines Bayesian Blocks to belong to a flare (shaded turquoise) if the flux level exceeds the mean plotted in pink. The flares of the half-clap method (shaded in pink and orange) are defined at their start and end by the minimum of half the block length and extrapolating the slope onto the valley block. (See Sec. 2.2 for further information.)
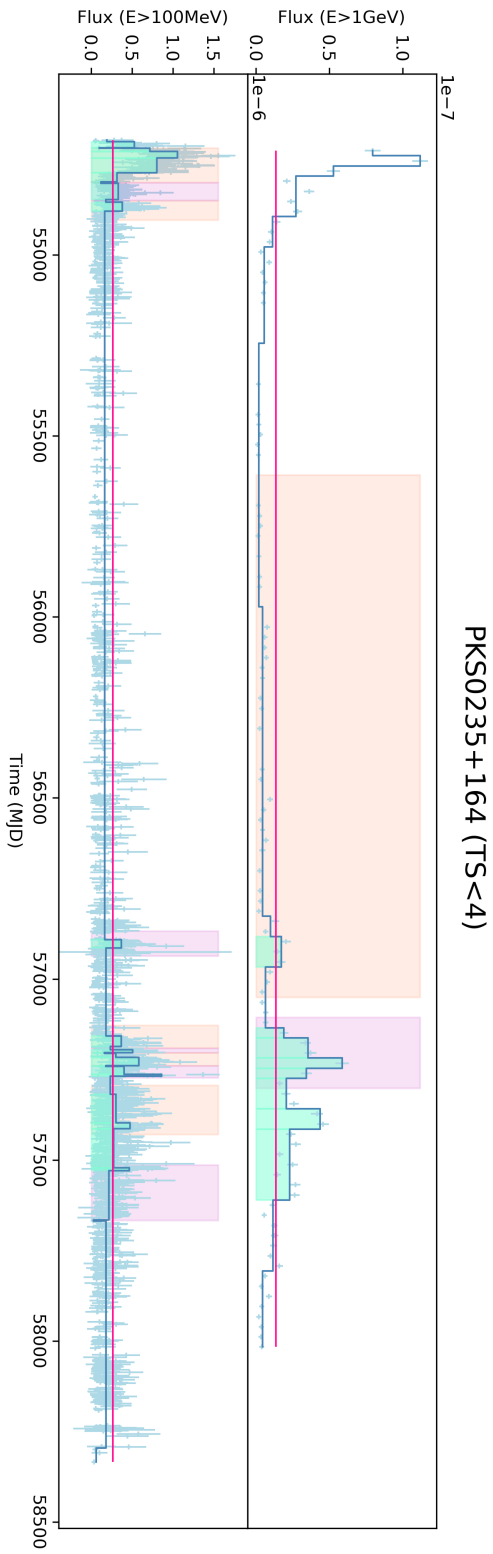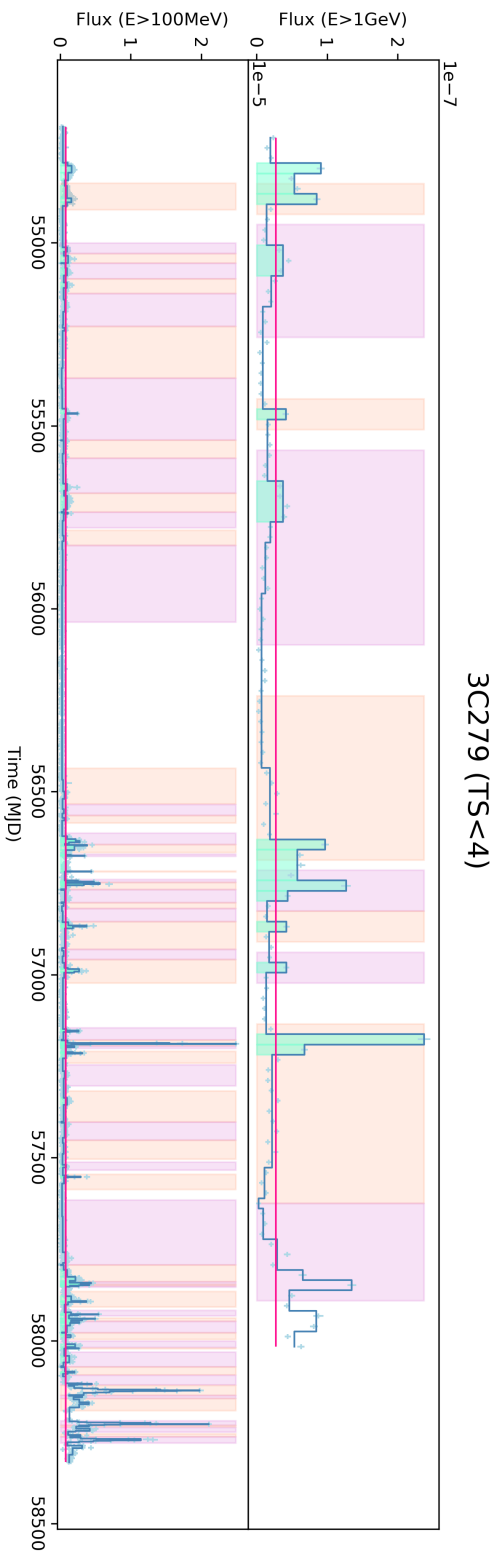
Figure A.3.: Light curve with observed flux of PKS 0235+164 in monthly (*top*) and daily (*bottom*) binning for energies over 1GeV and 100MeV, respectively. The baseline method defines Bayesian Blocks to belong to a flare (shaded turquoise) if the flux level exceeds the mean plotted in pink. The flares of the half-clap method (shaded in pink and orange) are defined at their start and end by the minimum of half the block length and extrapolating the slope onto the valley block. (See Sec. 2.2 for further information.)
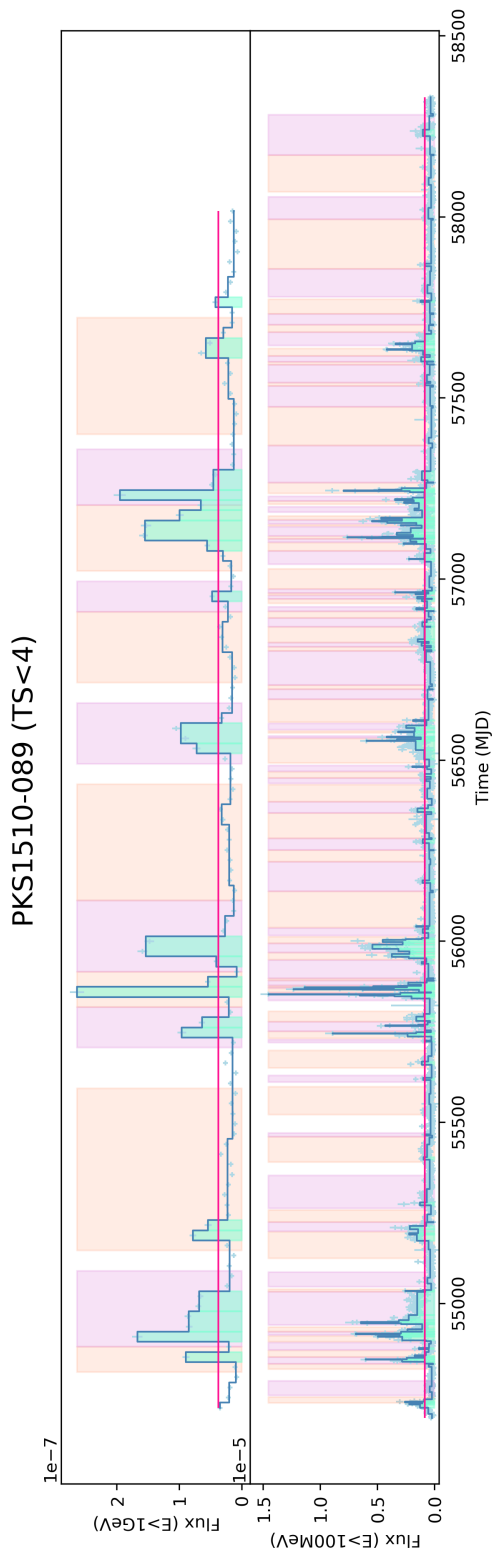
Figure A.4.: Light curve with observed flux of PKS 1510-089 in monthly (*top*) and daily (*bottom*) binning for energies over 1GeV and 100MeV, respectively. The baseline method defines Bayesian Blocks to belong to a flare (shaded turquoise) if the flux level exceeds the mean plotted in pink. The flares of the half-clap method (shaded in pink and orange) are defined at their start and end by the minimum of half the block length and extrapolating the slope onto the valley block. (See Sec. 2.2 for further information.)
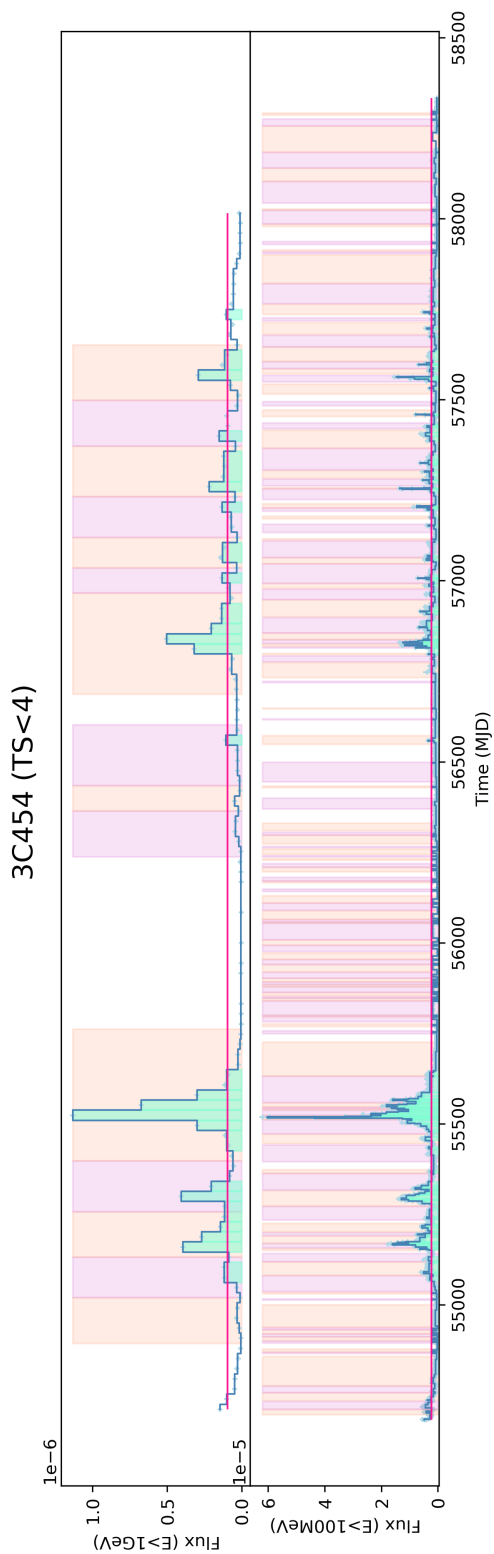
Figure A.5.: Light curve with observed flux of PKS 2155-304 in monthly (*top*) and daily (*bottom*) binning for energies over 1GeV and 100MeV, respectively. The baseline method defines Bayesian Blocks to belong to a flare (shaded turquoise) if the flux level exceeds the mean plotted in pink. The flares of the half-clap method (shaded in pink and orange) are defined at their start and end by the minimum of half the block length and extrapolating the slope onto the valley block. (See Sec. 2.2 for further information.)

Figure A.6.: Light curve with observed flux of 3C 273 in monthly (*top*) and daily (*bottom*) binning for energies over 1GeV and 100MeV, respectively. The baseline method defines Bayesian Blocks to belong to a flare (shaded turquoise) if the flux level exceeds the mean plotted in pink. The flares of the half-clap method (shaded in pink and orange) are defined at their start and end by the minimum of half the block length and extrapolating the slope onto the valley block. (See Sec. 2.2 for further information.)

Figure A.7.: Light curve with observed flux of 3C 279 in monthly (*top*) and daily (*bottom*) binning for energies over 1GeV and 100MeV, respectively. The baseline method defines Bayesian Blocks to belong to a flare (shaded turquoise) if the flux level exceeds the mean plotted in pink. The flares of the half-clap method (shaded in pink and orange) are defined at their start and end by the minimum of half the block length and extrapolating the slope onto the valley block. (See Sec. 2.2 for further information.)

Figure A.8.: Light curve with observed flux of 3C 454.3 in monthly (*top*) and daily (*bottom*) binning for energies over 1GeV and 100MeV, respectively. The baseline method defines Bayesian Blocks to belong to a flare (shaded turquoise) if the flux level exceeds the mean plotted in pink. The flares of the half-clap method (shaded in pink and orange) are defined at their start and end by the minimum of half the block length and extrapolating the slope onto the valley block. (See Sec. 2.2 for further information.)

Thank you!

Declaration

I, Sarah Maria Wagner, declare that this thesis titled, "Statistical properties of the high-energy gamma-ray flux variations of blazars" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.

- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.

- Where I have consulted the published work of others, this is always clearly attributed.

- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.

- I have acknowledged all main sources of help.

- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

27th of May 2020

Date, Signature